



**HAL**  
open science

# Développement des inférences pragmatiques et autisme : implicatures scalaires et métaphores à l'âge scolaire

Nicolas Petit

► **To cite this version:**

Nicolas Petit. Développement des inférences pragmatiques et autisme : implicatures scalaires et métaphores à l'âge scolaire. Sciences cognitives. Université Claude Bernard - Lyon I, 2023. Français. NNT : 2023LYO10246 . tel-04575648

**HAL Id: tel-04575648**

**<https://theses.hal.science/tel-04575648>**

Submitted on 15 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# THESE DE DOCTORAT DE L'UNIVERSITE CLAUDE BERNARD LYON 1

École Doctorale N° 479  
Neurosciences et Cognition (NSCo)

**Discipline** : Sciences cognitives

Soutenue publiquement le 22/11/2023, par :

**Nicolas Petit**

---

## Développement des inférences pragmatiques et autisme : implicatures scalaires et métaphores à l'âge scolaire

---

Devant le jury composé de :

<b>Stephanie DURRLEMAN</b> , Professeure, Swiss National Science Foundation, Université de Fribourg	Rapporteure
<b>Diana MAZZARELLA</b> , Professeure, Université de Neuchâtel	Rapporteure
<b>Maud CHAMPAGNE-LAVAU</b> , Directrice de recherche CNRS, Laboratoire Parole et Langage, Aix-en-Provence	Présidente
<b>Nicolas GEORGIEFF</b> , Professeur, Praticien hospitalier, Centre Hospitalier Le Vinatier, Université Claude Bernard Lyon 1	Examineur
<b>Jérôme PRADO</b> , Chargé de recherche CNRS, Centre de Recherche en Neurosciences de Lyon	Directeur de thèse
<b>Ira NOVECK</b> , Directeur de recherche CNRS, Laboratoire de Linguistique Française, Paris	Co-directeur de thèse
<b>Christina SCHMITZ</b> , Chargée de recherche CNRS, Centre de Recherche en Neurosciences de Lyon	Membre invitée



## Résumé

Les inférences pragmatiques permettent d'inférer ce qu'un locuteur *voulait dire* en disant ce qu'il a dit. Elles sont donc une compétence clef pour naviguer dans notre environnement social. Elles ont notamment une importance particulière dans le développement de l'enfant, et leur entrave dans des conditions comme l'autisme peut entraîner des conséquences négatives non-négligeables. Il est donc capital de bien comprendre les mécanismes à l'œuvre, et de bien les évaluer. Cette thèse se focalise sur deux types classiques d'inférences pragmatiques : les implicatures scalaires, dans lesquelles une expression peu informative implique la négation d'une alternative plus informative (par exemple « *quelques* » pour sous-entendre « *quelques, mais pas tous* ») et les métaphores (par exemple, « *Tu es la crème de mon café* »). Nous proposons deux nouvelles tâches expérimentales sur tablette visant à évaluer leur maturité à l'âge scolaire. Ces tâches ont dans un premier temps été déployées pour évaluer le développement typique de ces habiletés auprès d'un large échantillon d'environ 250 enfants neurotypiques. Alors que des travaux récents mettaient en évidence la précocité de certains précurseurs de ces inférences à l'âge préscolaire, nos résultats soutiennent que la fenêtre de 6 à 11 ans reste une importante période de développement de ces habiletés. Cette maturation est notamment visible pour les *implicatures scalaires*, puisque nous mettons en évidence dans une première étude que l'âge est associé à une propension grandissante à utiliser spontanément l'informativité pour inférer les intentions de communication d'un locuteur. De même, nous montrons dans une deuxième étude que les performances des enfants en *compréhension de références métaphoriques* progressent également fortement à cet âge. Nos résultats révèlent en outre une bascule dans les prédicteurs cognitifs associés à la compréhension des métaphores, de la théorie de l'esprit chez les enfants les plus jeunes aux compétences langagières formelles chez les plus âgés. Ces éléments suggèrent un changement qualitatif et pas seulement quantitatif dans le développement pragmatique à cette période. Fort de ces repères développementaux, nous avons utilisé ces deux tâches pour évaluer des enfants autistes dans une troisième étude. En effet, l'autisme est classiquement associé à une interprétation inhabituellement littérale du langage, mais la littérature expérimentale pousse à nuancer cette affirmation. En cohérence, nos données cliniques préliminaires confirment que le développement pragmatique à l'âge scolaire est également actif chez les enfants autistes, mais avec des trajectoires qui peuvent diverger de celles d'enfants neurotypiques. Plutôt que de révéler une incapacité pragmatique globale, ou un retard généralisé, nos données dévoilent un pattern spécifique associé à l'autisme, avec des atypies visibles ici sur les implicatures scalaires, mais pas sur les métaphores, à l'inverse de ce qui était attendu. Nous discutons la possibilité que le caractère spontané ou non d'une opération pragmatique soit un paramètre important à prendre en compte pour mieux comprendre les

particularités associées à l'autisme. Cette thèse développe plus généralement comment l'ensemble de ces résultats peuvent être utilisés pour interroger les modèles pragmatiques et les modèles de l'autisme, tout comme les perspectives théoriques et les implications pratiques qu'ils dessinent.

*Mots-clefs* : pragmatique expérimentale, autisme, développement du langage, cognition sociale, métaphore, implicatures scalaires, théorie de l'esprit

# Pragmatic inference development and autism: scalar implicatures and metaphors at school age

## Abstract

Pragmatic inferences are what allow us to infer what a speaker *meant* by saying what they said. As such, they represent an essential competence for navigating in social environments. They are of particular importance in children's development, and any impairment (e.g., in conditions such as autism) can bring about non-negligible consequences. It is thus crucial to get a clear understanding of the mechanisms underlying such skills, and to assess them properly. This thesis focuses on two classic examples of pragmatic inferences: scalar implicatures, by which a less informative expression can imply the negation of a more informative alternative (for instance "some" often implies "some but not all"), and metaphors (as in "You are the cream in my coffee"). We introduce two new tablet-based tasks to assess the development of these inferences at school-age. In a first step, these tasks were used to study typical development among a large sample of approximately 250 neurotypical children between 6 and 11 years old. While previous studies have shown early precursors of these inferences at pre-school age, our results demonstrate that school age remains a key developmental window. This development is visible with *scalar implicature*, as we show in a first study where age was associated with an increased tendency to spontaneously use informativity to infer speakers' meanings. Equally, we show in a second study that children's performance with *metaphoric references* also greatly improves during this developmental window. Furthermore, our results reveal a shift in the cognitive predictors associated with children's metaphor performance, transitioning from theory of mind in younger children, to formal language skills in older children. This suggests not only a quantitative improvement but also a qualitative change in metaphor comprehension during school age. Using these developmental reference points, the two tasks were used to assess autistic children in a third study. While autism is classically associated with an unusually literal interpretation of language, the experimental literature tends to nuance that claim. Consistently with this, our clinical data confirms that pragmatic inferences also progress in autistic children during school age, but with developmental trajectories that can diverge from neurotypical children. Rather than revealing any generalized lack of pragmatic capacity or general developmental delay, our data reveal a specific pattern of pragmatic performance in autism, with differences evident on scalar implicatures, but not on metaphors, in direct contrast to what was expected. We discuss the hypothesis that the spontaneousness of pragmatic operations might be an important factor to investigate when considering pragmatic abilities in autism. More generally, this thesis

expands on how those results can be used to examine pragmatic models and our understanding of autism, as well as the theoretical perspectives and the implications for clinical practice they convey.

*Key-words:* experimental pragmatics; autism; language development; social cognition; metaphor; scalar implicature; theory of mind



**LE VINATIER**  
PSYCHIATRIE UNIVERSITAIRE  
LYON MÉTROPOLE



Cette thèse a été préparée au *Centre Hospitalier Le Vinatier* (service SUNRISE, pôle de psychiatrie pour enfants et adolescents) et au *Centre de Recherche en Neurosciences de Lyon* (CRNL, équipe EDUWELL), situés au 95 boulevard Pinel à Bron (69 678).

- Ça s'est bien passé tes vacances ?
- Super bien, comme deux gouttes d'eau !

ÉCHANGE AVEC TOM (8 ANS)

Tu t'assoiras d'abord un peu loin de moi, comme ça, dans l'herbe. Je te regarderai du coin de l'œil et tu ne diras rien. Le langage est source de malentendus.

A. DE SAINT-EXUPERY (1943)

## Remerciements

Un léger vertige m'a saisi, à l'heure de présenter cette thèse, lorsque j'ai considéré ce qu'elle m'avait appris. Je n'en avais soupçonné ni l'étendue, ni la nature, et les chapitres à venir ne rendront compte que d'une partie de cet apprentissage. Il apparaîtra sans doute qu'en les écrivant j'ai tenté d'isoler des faits, des arguments, des raisonnements, ce qui pourrait donner à ce travail l'image d'une entreprise purement intellectuelle. Ce serait bien sûr une erreur. Car la première chose qu'il m'a permis de découvrir est que, si la science est une affaire d'idées et de démonstrations, tout cela n'existe que dans des corps qui vivent, respirent, ressentent, communiquent, et qu'on ne peut pas réellement la distinguer des émotions qui la nourrissent ou qui l'entravent, des gens et des relations qui la font vivre, des lieux qui l'accueillent, du matériel qui la rend palpable, et des innombrables épaules sur lesquelles elle se dresse. Il me paraît important de leur exprimer ici ma gratitude, donnant la mesure de l'aventure collective que cette thèse a aussi été, ce qui m'empêchera de formuler des remerciements aussi concis que voulu.

Sans surprise, mes premiers remerciements s'adressent à Ira Noveck, qui a accepté de diriger cette thèse lorsque le jeune orthophoniste que j'étais en 2017 lui a proposé ce projet ambitieux, peut-être trop ; je lui suis reconnaissant pour son exceptionnelle vision d'ensemble de la pragmatique expérimentale, qui a empreint tout ce travail, ainsi que pour ce que l'autonomie qu'il m'a laissée m'a aussi permis d'apprendre. Je remercie également Jérôme Prado qui a rejoint la direction de cette thèse un peu plus tard, m'ouvrant généreusement les portes de son équipe au CRNL où j'ai trouvé un environnement accueillant, et dont la considération m'a fait grandir. Je dois remercier également l'hôpital du Vinatier qui a hébergé ce projet, et plus particulièrement le Centre d'Évaluation et de Diagnostic de l'Autisme (CEDA), qui l'a vu naître puis grandir, dont les portes m'ont été ouvertes par Marie-Maude Geoffray en 2017. L'hôpital du Vinatier a par ailleurs permis ce travail en le finançant, aux côtés du ministère de la Santé, suite notamment aux encouragements de Sylvia Topouzkhania ; faire confiance à un orthophoniste pour mener un tel projet n'est pas (encore) si fréquent et je suis leur suis donc très reconnaissant.

Aussi passionnant ce travail ait été, j'ai d'abord été heureux d'arriver à mon bureau le matin pour retrouver mes collègues. Je remercie donc avec émotion l'équipe du CEDA, et en particulier Matias et Flavia bien sûr, Lucie, Hanène, Agathe, Marie-Luce, Sara, Saadia : leurs conseils, leurs compétences professionnelles et leur amitié m'ont beaucoup appris et aidé. Mener dans la durée ce projet a été autrement plus difficile que de l'initier, et je leur dois également. Au CRNL, je remercie notamment Léa et Justine, et tout particulièrement Cléa, finalement peu croisée au laboratoire mais beaucoup en dehors ; son écoute et ses nombreux conseils ont été déterminants. Je suis

également reconnaissant envers mes collègues de l'Université Claude Bernard Lyon 1, pour leur engagement et leur bienveillance, et tout particulièrement envers Floriane, dont le soutien a été essentiel pour parvenir au terme de ce travail. Naviguer entre des activités cliniques, scientifiques et universitaires a été un défi mais aussi, avec leur aide, une force.

Ce projet a pu se mettre en place grâce au support administratif de Lydie Sartelet à l'hôpital du Vinatier, grâce aux illustrations d'Étienne (Gol3m) qui ont fait du TIPi un outil attractif et représentatif d'une certaine diversité, grâce à la participation technique de PC Thomart du Studio ReRolled, grâce à Julie Penavayre qui a (entre autres) prêté sa voix pour certaines tâches du TIPi. La collecte des données a par ailleurs été permise grâce à la participation de nombreux partenaires : les écoles Saint-Denis, Saint-Sacrement, Immaculée Conception et Sainte-Marie à Lyon, l'académie de Limoges et l'école de Beynes à Égletons, la Maison de l'Enfance de Monplaisir, le centre CLEA coordonné par Charles-Edouard Fournier, les plateformes du service SUNRISE du Vinatier, dirigé par Marie-Maude Geoffray ; je suis très reconnaissant envers les équipes de ces structures qui m'ont accueilli et aidé en dépit d'un contexte pandémique. Pour ces expérimentations j'ai également bénéficié de l'aide d'Isabelle Petit, et de nombreuses étudiantes et étudiants en orthophonie accueillies en stage, ou qui m'ont fait confiance pour la direction de leur mémoire, dans le cadre du projet TIPi : merci à Sidonie Lévêque, Caroline Morrelet, Natacha Gentile, Emmanuelle Grisetti, Elodie Betin, Valentine Pernot, Mathilde Janier, Natalia Merida, Hugo Lamoitier. Merci plus généralement à l'ensemble des étudiant·es accueillis sur d'autres projets ou rencontrés en cours, qui participent d'une certaine façon à ma formation en me poussant à continuer d'apprendre.

Se prêter à une expérimentation scientifique et voir ses données personnelles enregistrées et utilisées par d'autres personnes, à des fins parfois difficiles à appréhender, n'est pas anodin. Aussi je tiens à remercier vivement les adultes et les nombreux enfants (et leurs familles) qui se sont prêtés aux expérimentations, et qui ont constitué la source même de ce travail. Leur confiance et leur enthousiasme m'obligent. Je remercie également les enfants, adolescents et familles rencontrés au CEDA, dont les expériences m'ont beaucoup appris, touché, et ont indéniablement nourri et motivé ce travail.

Une partie des analyses présentées dans cette thèse a été réalisée à l'IUSS de Pavia en Italie, où j'ai eu la chance d'être accueilli fin 2021 par Valentina Bambini que je remercie chaleureusement, de même que Luca Bischetti et Federico Frau, grâce à une bourse du réseau XPrag.it financée par l'ACRI. Le reste de ces travaux a (principalement) été réalisé à Lyon, tandis que l'écriture de cette thèse s'est (principalement) déroulée ailleurs, dans d'autres lieux où j'ai eu la chance d'être hébergé et inspiré. Je suis sûr qu'il y a dans ces lignes des échos de l'Estaque (merci à Caroline), des Alpes (merci à mes parents), de la Corrèze (merci à Aurore, à mes parents) et des bords de Saône (merci à Floriane). J'ai en outre eu l'opportunité

d'améliorer ce document grâce aux relectures avisées de Cléa, Floriane, Julie, ainsi que de Matias qui a également contribué à plusieurs articles issus de cette thèse. Merci.

Ce travail clôture enfin un long et passionnant parcours universitaire, initié en 2009, et je voudrais remercier les enseignantes et enseignants que j'ai rencontrés et qui m'ont aidé à grandir, dans les universités de Lyon 1, Toulouse 2 et Lyon 2. Je suis fier de me tenir désormais aux côtés de certains d'entre eux au département d'orthophonie de l'Université Claude Bernard. J'ai également une pensée pour les camarades avec lesquels j'ai avancé sur cette route et qui ont aussi contribué à ma formation, en particulier pour les Poux qui ont accompagné mes études d'orthophonie.

Au cours de ce travail, j'ai traversé des grands vides et des épiphanies, en oscillation entre l'isolement et la satisfaction, entre l'exaltation de mettre de l'ordre dans le chaos et l'intensité d'un sentiment de stupidité. J'ai longtemps cru que ces affects étaient *associés* à ce travail, mais il me semble désormais qu'ils le *constituent* aussi. J'ai beaucoup appris de ces mouvements également, ayant pu les apprivoiser grâce à des espaces où je pouvais respirer, m'exprimer, partager. En chantant, avec les amis et amies des Phonie Polies, Maude, le Ha'Chœur. En écrivant aussi, notamment avec le soutien des copains du collectif Pou et de Cécile. En échangeant avec Emmanuel Le Barbu. Et bien sûr grâce à la présence et aux attentions de mes proches, dont plusieurs ont déjà été cités ici. D'autres n'avaient rien à voir avec ce travail mais ont tout de même répondu présent, et leur soutien a compté, à un moment ou un autre : mille mercis en particulier à Julie, Nora, Anna, Aurore, Claire, la joyeuse bande corrézienne de Paris, mes parents, ma tante...

Je remercie aussi Jean-Baptiste Van der Henst et Anne Reboul d'avoir participé à mon comité individuel de suivi de thèse. Et finalement, puisque ce qui donne une valeur à un travail de recherche est son appréciation par des experts de la communauté scientifique, je remercie chaleureusement les deux rapportrices de cette thèse, Stephanie Durrleman et Diana Mazzarella, ainsi que les examinateurs, examinatrices ou membres invités du jury de soutenance, Maud Champagne-Lavau, Nicolas Georgieff et Christina Schmitz, pour l'honneur qu'ils me font de consacrer leur temps et leur attention à évaluer mon travail.

# Liste des publications

## 1. Publications incluses dans cette thèse

N. Petit, J. Prado, I. Noveck. Setting *some* tables: an action-based response time paradigm to study the development of scalar implicatures (*in preparation*). [Chapitre 4](#).

N. Petit, V. Bambini, L. Bischetti, J. Prado, I. Noveck. How do theory of mind and language skills impact metaphoric reference comprehension during children's school-age years (*under review* for the *Journal of Experimental Psychology: Learning, Memory and Cognition*). [Chapitre 5](#).

## 2. Publications incluses en annexes

N. Petit, I. Noveck, M. Baltazar, J. Prado. Assessing theory of mind in children: a tablet-based adaptation of a classic picture sequencing task (*under review* for *Child Psychiatry and Human Development*). [Annexe B](#).

Some (but not all) pragmatic inferences are difficult for autistic children (*in preparation*). [Annexe D](#).

# Sommaire

<b>SOMMAIRE</b> .....	13
<b>AVANT-PROPOS</b> .....	15
<b>INTRODUCTION GÉNÉRALE</b> .....	<b>17</b>
<b>CHAPITRE 1 • DIRE ET VOULOIR DIRE</b> .....	18
1. <i>Communiquer des intentions</i> .....	18
2. <i>De l'importance du sens littéral</i> .....	25
3. <i>Inférences pragmatiques et théorie de l'esprit</i> .....	38
<i>Vers une approche développementale</i> .....	50
<b>CHAPITRE 2 • LE DEVELOPPEMENT DES INFERENCES PRAGMATIQUES</b> .....	51
1. <i>Généralités</i> .....	51
2. <i>Développement des métaphores</i> .....	55
3. <i>Développement des implicatures scalaires</i> .....	63
<i>Développement typique et atypique</i> .....	69
<b>CHAPITRE 3 • INFERENCES PRAGMATIQUES ET AUTISME</b> .....	71
1. <i>Le second degré : au cœur de l'autisme ?</i> .....	72
2. <i>Métaphores et autisme</i> .....	80
3. <i>Implicatures scalaires et autisme</i> .....	88
4. <i>Compétence unique ou processus multiples</i> .....	94
<b>PROBLEMATIQUE DE LA THESE</b> .....	98
<b>CONTRIBUTIONS EXPÉRIMENTALES</b> .....	<b>101</b>
<b>CHAPITRE 4 • DEVELOPPEMENT TYPIQUE DES IMPLICATURES SCALAIRES (ETUDE 1)</b> .....	102
<i>Method</i> .....	108
<i>Results</i> .....	112
<i>Discussion</i> .....	115
<i>References</i> .....	120
<i>Point d'étape à l'issue de l'étude 1</i> .....	123
<b>CHAPITRE 5 • DEVELOPPEMENT TYPIQUE DES METAPHORES (ETUDE 2)</b> .....	124
<b><i>Experiment 1</i></b> .....	130
<i>Method</i> .....	131
<i>Results</i> .....	133
<i>Discussion</i> .....	134
<b><i>Experiment 2</i></b> .....	135
<i>Method</i> .....	136
<i>Results</i> .....	139
<i>Discussion</i> .....	145
<b><i>General Discussion</i></b> .....	146

<i>References</i> .....	150
<i>Point d'étape à l'issue de l'étude 2</i> .....	154
<b>CHAPITRE 6 • DONNEES CLINIQUES PRELIMINAIRES (ETUDE 3)</b> .....	155
<i>Introduction</i> .....	155
<i>Méthodologie</i> .....	156
<i>Résultats</i> .....	158
<i>Discussion</i> .....	163
<i>Conclusion et point d'étape à l'issue de l'étude 3</i> .....	168
<b>DISCUSSION GÉNÉRALE</b> .....	<b>170</b>
<b>CHAPITRE 7 • SYNTHÈSE ET DISCUSSION DES RESULTATS OBTENUS</b> .....	171
1. <i>Synthèse des résultats</i> .....	171
2. <i>Implications pour les modèles pragmatiques</i> .....	175
3. <i>Implications pour les modèles de l'autisme</i> .....	178
<b>CONCLUSION</b> .....	191
<b>RÉFÉRENCES</b> .....	<b>194</b>
<b>ANNEXES</b> .....	<b>223</b>
<b>ANNEXE A • MATERIEL SUPPLEMENTAIRE POUR L'ETUDE 1</b> .....	224
<b>ANNEXE B • ÉVALUER LA THEORIE DE L'ESPRIT CHEZ L'ENFANT (ETUDE ANNEXE)</b> .....	225
<b>Experiment 1</b> .....	228
<i>Methods</i> .....	229
<i>Results</i> .....	231
<i>Discussion</i> .....	233
<b>Experiment 2</b> .....	235
<i>Methods</i> .....	235
<i>Results</i> .....	237
<i>Discussion</i> .....	239
<b>General Discussion</b> .....	240
<i>References</i> .....	242
<i>Appendix: clinical cases</i> .....	247
<i>Supplementary materials</i> .....	249
<b>ANNEXE C • MATERIEL SUPPLEMENTAIRE POUR L'ETUDE 2</b> .....	255
<b>ANNEXE D • CERTAINES INFERENCES PRAGMATIQUES (MAIS PAS TOUTES) SONT DIFFICILES POUR LES ENFANTS AUTISTES (MANUSCRIT)</b> .....	261
<i>Methods</i> .....	264
<i>Results</i> .....	267
<i>Discussion</i> .....	271
<i>References</i> .....	276
<i>Supplementary materials</i> .....	280
<b>ANNEXE E • TABLE DES FIGURES</b> .....	283
<b>ANNEXE F • SERMENT DU DOCTEUR</b> .....	287

## Avant-propos

On pourrait être surpris d'entendre un adolescent répondre, alors qu'on lui demande s'il a un ou une petite amie : « *Oui, j'en ai plusieurs, mais aussi des grandes* ».

Cette réponse paraîtra toutefois moins étonnante si elle émerge d'un entretien conduit dans une unité d'évaluation et de diagnostic de l'autisme. En effet, l'autisme est typiquement associé à des difficultés à comprendre les sous-entendus, le second degré, le sens figuré, conduisant à une compréhension excessivement littérale du langage. Mettre en évidence cette particularité relève de la mission des orthophonistes de ces unités, qui cherchent à objectiver les compétences et les difficultés des jeunes accueillis dans les domaines de la communication et du langage. Ces évaluations doivent permettre d'identifier des entraves à leur participation sociale et à leur autodétermination, à les orienter vers des interventions et des aménagements qui limitent ces handicaps, et à étayer une démarche diagnostique qui participe à la construction identitaire des personnes et donne accès à un certain nombre de droits.

Des outils variés sont à disposition des cliniciens pour ce faire : tests cliniques standardisés et étalonnés, questionnaires parentaux, outils semi-structurés, observations cliniques mises en rapport avec des repères développementaux et de fonctionnement neurotypique. Si les tests standardisés permettent les observations les plus objectives et reproductibles, donc comparables à des normes fiables, il apparaît toutefois qu'il en existe peu qui soient satisfaisants dans le domaine de la communication sociale, intrinsèquement difficile à standardiser et pourtant crucial dans l'autisme. Au sein de l'équipe d'évaluation dans laquelle j'évoluais, j'ai par exemple proposé, le lendemain de mon entretien avec l'adolescent évoqué plus haut, un test normalisé de compréhension d'expressions figurées. Curieusement, le score qu'il a obtenu était supérieur à ce qui était attendu pour son âge, malgré les observations cliniques évidentes qui avaient été faites. Ce test montrait donc des limites flagrantes. Pour des enfants plus jeunes le besoin était plus évident encore, puisqu'aucun outil n'est même disponible.

C'est de ce constat qu'est né le projet TIPi, puis cette thèse. Il apparaissait en effet nécessaire d'amorcer la construction d'un nouvel outil standardisé pour capturer ce type de difficultés dans le quotidien, qui deviendrait un *Test des Inférences Pragmatiques informatisé* (TIPi). Conçu sur tablette tactile, le TIPi vise à évaluer une diversité de phénomènes langagiers implicites, en recueillant les réponses mais également les temps de réponses des enfants, et ce de façon contrôlée, automatisée et engageante. À travers la construction et la validation de cet outil, de nombreuses données seraient récoltées, auprès d'enfants neurotypiques ou non, qui pourraient également être exploitées pour mieux comprendre le développement de ces habiletés.

C'est ainsi que, dès le départ, le pari de ce projet de recherche est de concilier cette double motivation : d'une part le besoin appliqué d'un outil clinique, et d'autre part des questionnements théoriques, à l'intersection des sciences du langage, de la psychologie cognitive et du développement et de la psychiatrie de l'enfant et de l'adolescent, qui pourront secondairement profiter à la pratique clinique. Cette ambition qui peut sembler importante est finalement celle de l'orthophonie : faire l'interface entre la connaissance scientifique et un besoin social, à l'endroit de la santé de la communication et du langage.

Cette thèse développe certains de ces questionnements théoriques. Afin d'en restreindre le périmètre, nous nous focaliserons sur la période développementale de l'âge scolaire (6-11 ans), et sur deux types (emblématiques) d'inférences pragmatiques : les métaphores et les implicatures scalaires. Dans une première partie théorique, nous établirons un cadre de référence et évoquerons sans viser l'exhaustivité des repères issus de travaux réalisés chez l'adulte, qui pourront être mobilisés dans la suite de ce travail (*chapitre 1*). Puis nous examinerons l'état des connaissances quant à *quand* et *comment* se développent typiquement les inférences pragmatiques chez l'enfant (*chapitre 2*). Un dernier chapitre théorique portera sur les inférences pragmatiques dans l'autisme, qui constitue à la fois un besoin clinique important et un prisme pour l'étude scientifique des phénomènes qui nous intéressent (*chapitre 3*). Nous présenterons ensuite les trois contributions expérimentales de cette thèse, avec un premier article sur le développement typique des implicatures scalaires (*chapitre 4*) et un deuxième article sur les prédicteurs cognitifs de la compréhension des métaphores dans le développement typique de l'enfant (*chapitre 5*). Nous rapporterons ensuite des données cliniques préliminaires sur la compréhension de ces phénomènes, implicatures scalaires et métaphores, chez des enfants autistes (*chapitre 6*). Nous pourrions finalement examiner dans une discussion générale (*chapitre 7*) les apports de ces données par rapport aux questions de recherches, discuter leurs implications théoriques comme cliniques, et les perspectives qu'elles soulèvent.<sup>1</sup>

---

<sup>1</sup> Tout au long de cette thèse, les notes de bas-de-page seront principalement réservées à des précisions d'ordre terminologique ou linguistique, qu'un lecteur intéressé par ces questions pourra consulter mais dont la lecture n'est pas nécessaire pour la compréhension du propos.

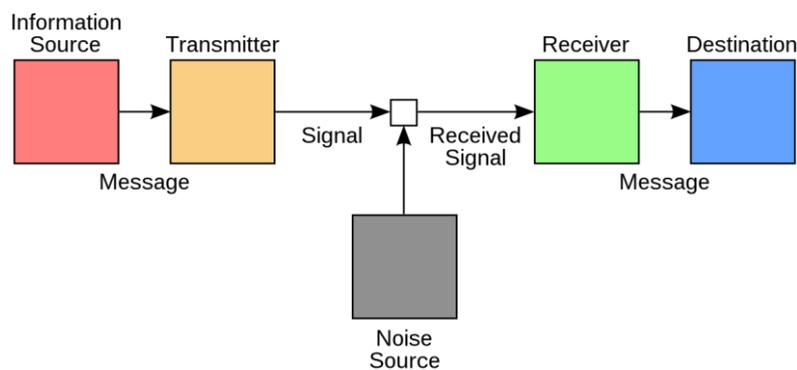
# INTRODUCTION GÉNÉRALE

# Chapitre 1 • Dire et vouloir dire

## 1. Communiquer des intentions

### 1.1. Deux modèles de la communication

La communication peut spontanément être décrite comme un processus de transmission d'information entre un émetteur et un récepteur. Un message est encodé au niveau d'un émetteur, puis transmis via un certain canal, et enfin décodé au niveau d'un récepteur. L'utilisation d'algorithmes associatifs identiques à l'émission et à la réception permet de rétablir la forme originale du message à l'issue du processus, de sorte que l'information a été *transférée*. On pourrait ainsi transporter de l'information (par exemple dans un livre) comme on transporterait des biens (par exemple dans un container). Ce *modèle du code* a été formalisé par Shannon en 1948 (figure 1-1), précisant que le transport de l'information est susceptible d'être perturbé par des erreurs, du bruit, et qu'un système de communication efficace doit pouvoir gérer ces problèmes. Très intuitive, cette modélisation de la communication, « *mère de tous les modèles* » (Hollnagel & Woods, 2005), a eu une influence considérable dans les sciences de l'ingénierie, de l'information et de la communication, et plus largement dans les sciences humaines et sociales, faisant l'objet de nombreuses complexifications. Pour ce qui est de la communication humaine, il a par exemple été utilisé pour expliquer la redondance importante de l'information dans le langage, comme un moyen d'assurer l'efficacité de la communication, malgré la présence de bruit.



**Figure 1-1 :** *Modèle du code (Shannon, 1948)*

Le modèle du code peut être appliqué à de nombreux systèmes, et notamment à la plupart voire toutes les formes de communication animale non-humaine (Scott-Phillips, 2014). Cependant, s'il peut décrire quelques comportements de communication humains, tels que le rire involontaire, il rencontre des difficultés

insurmontables à modéliser la communication linguistique entre humains. Prenons le cas de l'échange suivant :

*A : On pourrait prendre un verre ce soir ?*

*B : Je dois travailler sur mon article...*

Le modèle du code échoue fondamentalement à expliquer que l'énoncé « *Je dois travailler sur mon article...* » puisse signifier « *Non* », ou encore « *Non, je ne suis pas disponible* » dans ce contexte précis (mais pas dans d'autres). Un paradigme radicalement différent est nécessaire pour modéliser ce type d'usages. C'est ce que propose Paul Grice (1969, 1989), en introduisant une distinction fondamentale entre ce que *dit* un locuteur de ce qu'il *veut dire*<sup>2</sup> en disant cela, c'est-à-dire ses intentions de communication. La distance entre ce qui est dit et l'intention de communication peut être évidente, comme dans l'exemple précédent, ou plus restreinte, mais elle est inhérente à l'utilisation humaine du langage. Tout énoncé peut en effet véhiculer plusieurs intentions de communication, selon le contexte. La simple phrase « *Il fait nuit* » peut ainsi être utilisée comme l'expression littérale d'une information sur la luminosité, comme une requête indirecte d'allumer la lumière, comme le refus d'une proposition de sortir prendre l'air, etc. Le sens linguistique d'un énoncé sous-détermine toujours les intentions de communication du locuteur (Carston, 2008). Il ne peut donc exister de code suffisant et en cela, par essence, *le langage est source de malentendus*.

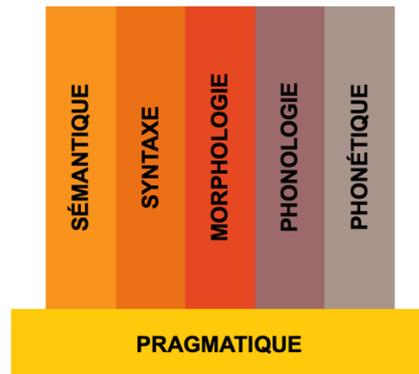
Dans cette nouvelle perspective, la communication est alors définie comme l'expression et la reconnaissance d'intentions<sup>3</sup> (Grice, 1989). On qualifie de *pragmatiques* ces habiletés et leur étude scientifique (Y. Huang, 2017; Scott-Phillips, 2017). Si ce qui est dit peut être linguistiquement décodé, ce n'est pas le cas des intentions de communication du locuteur. Il n'y a donc pas à proprement parler de transfert d'information, mais un locuteur qui donne des indices de ses intentions (ostension) ; ces indices peuvent être utilisés par un autre individu pour interpréter, plus ou moins approximativement, les intentions du locuteur (inférence) : c'est un modèle *ostensif-inférentiel*. Ce type de communication n'est donc pas nécessairement verbal, pouvant s'appuyer sur des comportements tels que les gestes, les mimiques, mais il exploite la richesse du système conventionnel que constitue une langue, qui permet de fournir des indices particulièrement précis (c'est la '*pragmatique au sens fort*' selon Scott-Phillips, 2017, voir figure 1-2). À la différence d'un container, un livre ne contient donc pas d'information en lui-même ; Sperber et Wilson (1989) l'expliquent ainsi : « *nous n'avons pas littéralement mis nos pensées sur du papier. Ce*

---

<sup>2</sup> Originellement en anglais : l'opposition entre *sentence meaning* et *speaker's meaning*.

<sup>3</sup> Plus précisément, dans les travaux de Grice ces intentions recouvrent à la fois l'intention de montrer que nous cherchons à communiquer (intention communicative) et l'intention de montrer *ce* que nous voulons communiquer (intention informative). Dans ce travail, nous n'exploiterons pas cette distinction et parlerons plus globalement d'*intentions de communication*.

que nous avons mis sur le papier, ce sont de petites marques noires, dont vous avez en ce moment une copie sous les yeux. Quant à nos pensées, elles sont restées là où elles ont toujours été : dans nos cerveaux » (p. 11).



**Figure 1-2 :** Représentation de la pragmatique au sens fort, comme base cognitive du langage (reproduction et traduction de Scott-Phillips, 2017)

La pragmatique s'est ainsi attachée à décrire les relations qu'entretiennent les énoncés avec les intentions qui les sous-tendent. Ayant introduit cette distinction, Grice propose que ces rapports au sein d'une conversation soient rationnels, et régis par un *principe de coopération*, en vertu duquel chaque interlocuteur s'engage à coopérer pour faire avancer la conversation et à « [contribuer] de la façon requise, au moment où cela se produit, par le but ou la direction acceptée de l'échange dans lequel [il est engagé] » (Grice, 1975, p. 45, traduction personnelle). Ce principe est décliné en quatre maximes conversationnelles :

1. *Maxime de qualité* : Ne dites pas ce que vous croyez être faux, ni ce que vous n'avez pas de raisons suffisantes de croire vrai
2. *Maxime de quantité* : Contribuez de façon aussi informative que nécessaire pour le but de cet échange, mais pas de façon plus informative que nécessaire
3. *Maxime de relation* : Soyez pertinent
4. *Maxime de manière* : Évitez d'être obscur, ambigu, soyez bref et ordonné

L'inférence pragmatique des intentions de communication du locuteur peut alors être opérée dans l'interaction entre ces maximes et le sens linguistiquement encodé de l'énoncé. La réponse de l'exemple précédent (« *Je dois travailler sur mon article...* ») enfreint manifestement la maxime de relation, puisqu'elle ne répond pas à la question qui est posée et n'est donc, littéralement, pas pertinente. Mais partir du principe que ce locuteur est coopératif et respecte la maxime de relation permet d'établir que ce qu'il *veut dire* est pertinent, et ainsi d'inférer que son intention est bien de répondre à la question qui lui est posée, et donc probablement de refuser. La maxime de relation a délibérément été bafouée, et ainsi *exploitée* pour communiquer implicitement une intention : on parle alors d'*implicature*.

Ces maximes peuvent être utilisées pour décrire d'autres types d'implicatures, comme les métaphores, qui exploiteraient la maxime de qualité. Grice propose l'exemple de l'énoncé suivant : *Tu es la crème de mon café* (Grice, 1989, p. 34). Cet énoncé, littéralement impossible donc faux, enfreint la maxime de qualité, poussant l'auditeur à rechercher une intention de communication alternative qui respecterait la maxime, et à inférer la signification de cette métaphore. Ces principes peuvent donc être utilisés pour adapter une signification littéralement fautive, mais également pour préciser et rendre plus informative une expression qui serait vraie. C'est le cas de l'énoncé *Quelques étudiants ont validé leur examen*, par exemple. Bien que cet énoncé soit, d'un point de vue sémantique, compatible avec le fait que tous les étudiants aient réussi leur examen, il serait alors plus informatif d'utiliser le terme *tous*, plutôt que *quelques*. L'exploitation de la maxime de quantité (contribuez de façon aussi informative que nécessaire) permet donc d'établir que tous les étudiants n'ont pas validé leur examen, et que certains ont échoué. En d'autres termes, 'quelques' est pragmatiquement enrichi pour signifier implicitement 'quelques, mais pas tous'. On désigne cette utilisation d'un terme moins informatif comme la négation d'un terme plus informatif au sein d'une *échelle* (ici, l'échelle <quelques, tous>) par le terme d'*implicature scalaire* (Horn, 1989). À noter que dans cet exemple, un professeur aurait pu annoncer « *quelques étudiants ont validé leur examen* » avant de révéler qu'aucun étudiant n'avait en réalité échoué, et se défendre d'avoir menti. Il aurait, tout au plus, joué sur les mots. C'est en effet une propriété importante et caractéristique de l'implicite véhiculé par les implicatures : il peut être *annulé*, ou corrigé par le contexte. Dans le premier exemple cité (l'invitation à prendre un verre), l'interlocuteur aurait ainsi pu compléter sa réponse en disant « Je dois travailler sur mon article... mais d'accord, je peux bien prendre un peu de temps pour moi », sans pour autant se dédire. Cette caractéristique distingue les implicatures des présuppositions (l'énoncé « Pierre a arrêté de fumer » présuppose que précédemment, Pierre fumait) et des implications (l'énoncé « le président a été assassiné » implique « le président est mort »), qui elles ne sont pas annulables.

En tant qu'exemples distincts mais prototypiques d'implicatures, les implicatures scalaires et les métaphores ont fait l'objet d'une littérature particulièrement fournie et seront au cœur de ce travail de thèse.

## **1.2. La puissance de la communication ostensive-inférentielle**

La communication ostensive-inférentielle, permettant une utilisation indirecte du langage, peut donc apparaître extrêmement complexe, cognitivement coûteuse et génératrice d'ambiguïtés et donc de problèmes de communication, en comparaison de l'utilisation d'un code, même dans le cas d'un système conventionnel complexe comme le langage humain. Les ambiguïtés du discours sont d'ailleurs une cause majeure de conflits interpersonnels (Tannen, 1991). Mais la communication ostensive-inférentielle présente un avantage économique important puisqu'elle permet de communiquer beaucoup en disant peu, exploitant ainsi la capacité relativement limitée

du langage parlé humain pour transmettre de l'information (capacité estimée à environ 39 bits/s de façon translinguistique [Coupé et al., 2019], soit environ trois millions de fois moins que la capacité moyenne d'une connexion internet en France [Cable.co.uk, 2022]). De plus, en exploitant la capacité à raisonner sur les états mentaux d'autrui, elle peut s'affranchir des associations préexistantes qui sont nécessaires aux codes. Elle permet donc une utilisation inédite et créative de n'importe quel comportement, dont le langage, pour communiquer de façon riche et flexible. Au cours de l'évolution, l'émergence de ce type de communication aurait ainsi conféré un avantage adaptatif important à l'espèce humaine, qui navigue dans de larges groupes sociaux. Une capacité à manipuler les états mentaux d'autrui (sur le versant expressif) ou à inférer les représentations d'autres individus (sur le versant réceptif) constitue en effet dans ce cadre un avantage certain, de façon générale, qui peut se déployer dans différents domaines importants tels que la cohésion d'un groupe, la chasse, la séduction, etc. (Scott-Phillips, 2014). Cette habileté humaine pourrait même avoir permis l'émergence des langues et du langage (Scott-Phillips, 2014), bien que ce point fasse l'objet de controverses (voir par exemple Durrleman et al. [2022], ou Reboul [2015], qui défend l'hypothèse de l'émergence du langage comme un outil cognitif de pensée, secondairement utilisé à des fins communicatives).

Plus précisément, l'utilité sociale d'une communication verbale implicite a par ailleurs été décrite dans les théories de la politesse (Brown & Levinson, 1987; H. H. Clark, 1996; Kerbrat-Orecchioni, 1998), qui proposent que les formulations indirectes servent de « lubrifiant social », permettant de maintenir la coopération et l'équilibre de l'échange, et de préserver la face de chacun des interlocuteurs, au même titre que des marques de sympathie (e.g. compliments), ou de déférence (e.g. excuses). Typiquement, un contenu négatif (par exemple, critiquer le jeu d'un partenaire lors d'une partie de tennis) peut être exprimé dans une forme positive (« Quel beau jeu... »<sup>4</sup>), perçue comme moins menaçante. Dans le cas plus spécifique des implicatures scalaires, une proposition menaçante communiquée de façon implicite par un énoncé sous-informatif, est ainsi perçue comme plus polie et gentille (Bonnefon et al., 2009), et traitée avec cette réserve par l'auditeur (Mazzarella, 2015; Mazzarella, Trouche, et al., 2018). Une proposition communiquée à l'aide d'une implicature scalaire, plutôt qu'explicitement, permet au locuteur de s'en distancer, et limite ainsi le risque pour sa réputation associé au fait qu'elle puisse se révéler fausse (Mazzarella, Reinecke, et al., 2018). En d'autres termes, ces implicatures sont également utilisées comme un outil pour moduler notre engagement envers le contenu véhiculé, et ainsi stabiliser la communication en modulant le degré de confiance qui peut être accordée à chaque énoncé.

Au-delà du maintien de la coopération, la *théorie du locuteur stratégique* de Pinker (J. J. Lee & Pinker, 2010; Pinker, 2007) modélise la communication indirecte comme un outil permettant de gérer la possibilité que la relation ne soit *pas* coopérative, et les

---

<sup>4</sup> Exemple adapté de Dews et al. (1995)

risques que cela fait encourir. L'exemple prototypique avancé est celui d'une tentative de corruption d'un policier après une contravention. Dans cette situation, un automobiliste qui tenterait de corrompre le policier, prendrait un risque en ignorant à l'avance si celui-ci est corruptible ou honnête. Si le policier est corruptible, alors la contravention sera évitée au prix d'un pot-de-vin, et le coût sera réduit, mais si le policier est honnête, le coût sera majoré car la contravention sera doublée d'une autre amende pour tentative de corruption. Dans cette situation, un locuteur stratégique réduirait les risques qu'il prendrait en communiquant sa proposition de façon implicite plutôt que directe. En effet, comme discuté plus haut, le contenu implicite est annulable, et donc niable : il est toujours possible de démentir la tentative de corruption. En proposant un pot-de-vin de façon implicite, en disant par exemple « *Mince, la meilleure solution est peut-être de régler la contravention ici, pour éviter toute la paperasse administrative ?* » (J. J. Lee & Pinker, 2010), un locuteur stratégique en tire les bénéfices si le policier est corruptible, et limite le coût à la contravention initiale, si le policier est honnête. La situation du pot de vin est expérimentalement idéale car les coûts et les bénéfices peuvent être directement quantifiés, mais Pinker étend cette théorie à toutes les situations où le coût n'est pas financier mais émotionnel, et où la sanction n'est pas judiciaire mais sociale ; c'est le cas de la menace, de la séduction ou de la sollicitation d'un don, situations dans lesquelles des formulations indirectes et implicites sont particulièrement fréquentes. L'exemple évoqué précédemment peut ainsi être éclairé sous un nouveau jour :

*A : On pourrait prendre un verre ce soir ?*

*B : Je dois travailler sur mon article...*

Ici, la proposition de A pourrait tout à fait être une tentative de séduction, sous-entendant un intérêt sentimental ou sexuel, mais sous la forme d'une proposition de boire un verre, le protégeant de l'éventualité que B refuse une telle évolution de leur relation. Réciproquement, la réponse de B peut être une façon de refuser implicitement cette proposition en préservant la coopération et en préservant la face de A qui serait menacée par une réponse directe comme « Non ».

L'implicite est ainsi omniprésent dans la communication humaine. Le seul cas de l'ironie a été observé dans 7 % à 8 % des énoncés d'authentiques conversations entre jeunes adultes ou entre amis (Gibbs, 2000; Tannen, 1984), soit en moyenne une occurrence toutes les 2 minutes de conversation (Gibbs, 2000). Ce type d'énoncés est même encore plus fréquemment observé dans des échanges par messagerie instantanée que dans des interactions en face à face (Hancock, 2004), et présent dans la quasi-totalité d'e-mails échangés entre amis, même sur des messages courts (Whalen et al., 2009). Les métaphores sont quant à elles considérées comme un outil particulièrement puissant pour transmettre des idées complexes, abstraites, ou personnelles (Gibbs, 2008; Lakoff & Johnson, 1980). Elle rendent ainsi la communication plus efficace pour séduire (Gao et al., 2017) ou pour convaincre (Sopory & Dillard, 2002). De plus, utiliser des métaphores nécessite l'activation de connaissances sociales partagées, qui

augmente la sensibilité des interlocuteurs aux stimuli sociaux et établit ainsi une intimité entre eux (Bowes & Katz, 2015; Horton, 2007). Elles sont donc fréquemment observées dans les conversations (Steen et al., 2010), et plus particulièrement encore dans un contexte éducatif (Littlemore, 2016; Steen et al., 2010), ou en psychothérapie (Mathieson et al., 2016; Stoddard & Afari, 2014).

Par conséquent, les inférences qui donnent accès à ces phénomènes ont une importance considérable dans la vie des individus. Dans l'enfance et l'adolescence, où elles se développent particulièrement, de bonnes habiletés pragmatiques en général sont ainsi associées à une plus grande popularité parmi les pairs, et à de moindres difficultés socio-émotionnelles et comportementales (Gottman et al., 1975; Helland et al., 2014; Murphy et al., 2014; Petranovich et al., 2017; Wolters et al., 2014). Dans une étude longitudinale de grande ampleur, Law et al. (2015) ont par exemple montré que l'association entre un désavantage social à la naissance et des difficultés de comportements à l'adolescence était médiée par une mesure des compétences pragmatiques à l'âge scolaire, à 9 ans. En ce qui concerne la compréhension des métaphores plus particulièrement, Del Sette et al. (2021) ont évalué la relation de cette compétence avec la popularité auprès des pairs de 126 enfants à deux reprises, à un an d'intervalle, à environ 10 et 11 ans. Les résultats de cette étude longitudinale ont mis en évidence un lien bidirectionnel négatif entre la compréhension des métaphores et le rejet par les pairs : les enfants moins rejetés à 10 ans montraient une meilleure compréhension des métaphores un an plus tard, et les enfants avec une meilleure compréhension des métaphores initialement étaient moins rejetés que les autres à 11 ans. Dans les populations cliniques, des troubles pragmatiques sont associés à des difficultés notamment sociales (Miranda et al., 2020) et sont prédictrices de difficultés relationnelles ultérieures (Whitehouse et al., 2009). Cette association est vraisemblablement causale, et pas seulement corrélatoire, puisque des interventions ciblées sur les habiletés pragmatiques ont montré qu'elles pouvaient améliorer la qualité de vie générale de patients avec de telles difficultés (Bambini et al., 2022). L'étude du développement des inférences pragmatiques représente donc un enjeu important, et nous nous y consacrerons au [chapitre 2](#) pour ce qui est du développement typique et au [chapitre 3](#) pour le cas de l'autisme.

### **1.3. Différents modèles pragmatiques**

Suite à la distinction introduite par Grice entre le sens d'un énoncé et les intentions de communication du locuteur, différents modèles ont été développés pour expliquer l'articulation entre ces types de significations. La proposition théorique de Grice du principe de coopération et des maximes, visant à décrire ce que font les locuteurs, et pourquoi (niveau dit computationnel par Marr, 1982), a été transcrite à un niveau algorithmique (*comment* ils le font) en un modèle dit pragmatique standard (Standard Pragmatic Model, SPM). Dans ce modèle trois étapes sont nécessaires pour comprendre une implicature et inférer l'intention de communication d'un locuteur :

- (1) Décoder le sens de l'énoncé
- (2) Repérer qu'une maxime a été enfreinte
- (3) Rechercher une intention de communication respectant le principe de coopération

Ce modèle cependant a fait l'objet de critiques, et plusieurs alternatives ont été proposées. Ces alternatives se sont particulièrement développées en s'appuyant sur des techniques expérimentales permettant de les éprouver au regard de la réalité psychologique du fonctionnement humain, s'extrayant alors d'une appartenance exclusive aux domaines de la philosophie du langage et de la linguistique. On parle dès lors de *pragmatique expérimentale* (Noveck, 2018). Sans viser l'exhaustivité, nous évoquerons dans la suite de ce chapitre des modèles et travaux utiles pour cerner les enjeux de cette thèse. Certains ciblent plus particulièrement les implicatures scalaires ou les métaphores, et se focalisent notamment sur la place du sens linguistiquement encodé de l'énoncé (*partie 2*). D'autres sont plus généraux et se concentrent sur la place du raisonnement sur les états mentaux d'autrui dans les processus pragmatiques (*partie 3*).

## 2. De l'importance du sens littéral

Plusieurs modèles ont été développées suite aux travaux de Grice, dont certains seront évoquées ici dans la mesure où ils offrent l'opportunité d'étudier les rapports entre le sens littéral et le sens pragmatique d'une expression (les intentions de communication). L'abord expérimental de cette question a souvent été d'évaluer le *coût* de la compréhension pragmatique d'un énoncé, par rapport à son décodage littéral. Nous nous focaliserons ici sur les deux phénomènes au cœur de cette thèse : les implicatures scalaires et les métaphores, en évoquant des expérimentations réalisées auprès d'adultes ; ces repères qui nous permettront ensuite de nous focaliser sur les travaux développementaux au chapitre suivant (*chapitre 2*).

### 2.1. Implicatures scalaires

La proposition de Grice a particulièrement été développée par Horn (1989) qui est à l'origine de l'expression d'*implicature scalaire*, formalisant que ce type d'implicatures reposerait sur des échelles qui seraient intégrées au système de la langue. Outre l'échelle <quelques, la plupart, tous> évoquée ci-dessus, la langue comporterait ainsi de nombreuses échelles telles que <ou, et>, <pouvoir, devoir>, ou <sympa, génial>, de sorte que la disjonction dans « J'apporterai du vin *ou* des chips » exclue implicitement la conjonction (je n'apporterai pas à la fois du vin *et* des chips). Ou encore que le verbe de possibilité dans « Tu *peux* porter une cravate » exclue implicitement la nécessité (tu n'es pas *tenu* de porter une cravate), ou qu'un qualificatif modéré dans « Le film était *sympa* » exclue implicitement un qualificatif plus fort (le

film n'était pas *génial*), etc. Prolongeant cette approche dite neo-gricéenne dans sa théorie des Implicatures Conversationnelles Généralisées, Levinson (2000) propose que ces implicatures soient réalisées *par défaut*, selon un principe (parmi plusieurs) basé sur les maximes gricéennes : ce qui n'est pas dit n'est pas le cas (Q-Principe). Ainsi '*quelques*' serait automatiquement interprété comme '*quelques-mais-pas-tous*', grâce à l'application de cette heuristique, interprétation qui pourrait être révisée dans des cas particuliers, si le contexte l'exige. Cette proposition présente donc une différence importante avec le modèle SPM tiré directement des travaux de Grice, puisque dans le SPM l'interprétation pragmatique d'un terme scalaire est le fruit d'un travail d'enrichissement à partir de son sens littéral, donc cognitivement coûteux, tandis qu'elle est automatique et non-coûteuse pour Levinson. Cette opposition a donné lieu une littérature expérimentale fournie dans le but d'examiner si l'implicature scalaire était le fruit d'un effort cognitif par rapport à une lecture littérale. Outre les travaux développementaux (notamment l'étude de Noveck [2001], voir [chapitre 2](#)), nous détaillerons ici des études auprès d'adultes exploitant les temps de réponse ou de lecture, l'induction expérimentale d'une surcharge cognitive et des mesures neurophysiologiques, qui renseignent sur le coût cognitif des implicatures scalaires.

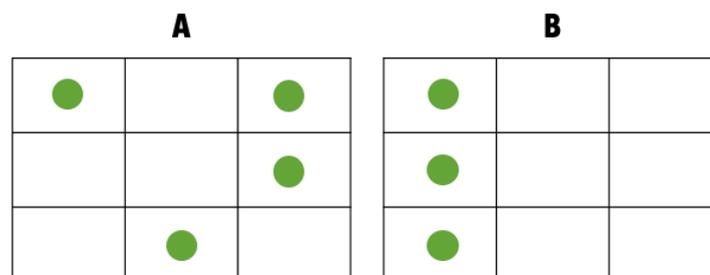
Le paradigme expérimental classiquement utilisé pour évaluer les implicatures scalaires consiste à présenter à des participants des phrases incluant de telles expressions et à leur demander d'évaluer si ces phrases sont vraies ou correctes (paradigme de vérification de phrases). Par exemple la phrase « quelques mammifères sont des éléphants » devrait facilement être considérée correcte, tandis que la phrase « quelques éléphants sont des insectes » devrait facilement être jugée fautive. En revanche, la phrase sous-informative « quelques éléphants sont des mammifères » est jugée différemment selon si le quantifieur est interprété de façon littérale (quelques-et-possiblement-tous, l'énoncé est alors correct) ou de façon pragmatique (quelques-mais-pas tous, l'énoncé est alors incorrect). C'est le type de phrases (écrites) que Bott et Noveck (2004) ont utilisé dans une série de 4 expériences, associés à d'autres conditions contrôles équivalentes utilisant le quantifieur 'tous'. Dans les deux premières expériences, il était explicitement demandé aux participants d'interpréter 'quelques' de façon pragmatique ou logique<sup>5</sup>. Les résultats montrent que les adultes commettent plus d'erreurs et répondent plus lentement lorsqu'on leur impose une interprétation pragmatique que lorsqu'on leur impose une interprétation logique. Dans une troisième expérience, les participants sont laissés libres de leur interprétation, ce jugement étant utilisé pour déterminer si leur interprétation est spontanément pragmatique ou logique. Alors que les participants répondent sans difficultés à tous les items contrôles, avec un taux de réponses correctes d'environ 92%, ils se montrent beaucoup plus équivoque dans la condition critique, où 41 % des réponses traduisent des interprétations logiques, contre 59% d'interprétations pragmatiques. Crucialement, la comparaison des temps de réponses pour ces deux types

---

<sup>5</sup> Nous reprenons ici le terme d'interprétation 'logique' employé par les auteurs mais aurions pu évoquer de façon équivalente une interprétation 'littérale' ou 'sémantique'.

d'interprétation révèle que les interprétations pragmatiques prennent plus de temps que les interprétations logiques, mais également que les réponses aux phrases contrôles. Cet effet a été récemment répliqué à une large échelle, dans une implémentation web de cette tâche (Ronderos & Noveck, 2023). Dans une quatrième expérience enfin, Bott et Noveck (2004) ont modifié le paradigme afin d'appliquer une pression temporelle sur certains participants, visant à réduire la disponibilité de leurs ressources cognitives. Si certains participants continuent de disposer d'un temps confortable pour répondre (3000 ms), d'autres se voient imposer une contrainte plus forte (900 ms). Alors que cette contrainte a peu d'impact sur toutes les conditions contrôles, imposer une contrainte temporelle fait chuter le taux de réponses pragmatiques dans la condition critique de 44 % à 28 %. L'ensemble de ces résultats convergent donc vers l'hypothèse que les implicatures scalaires soient dérivées au prix d'un certain effort cognitif, puisque i) il est plus difficile d'imposer à des participants une réponse pragmatique qu'une réponse logique, ii) qu'une interprétation pragmatique prend plus de temps qu'une réponse logique lorsque les participants sont laissés libres de leur interprétation et iii) que limiter les ressources cognitives des participants en imposant une contrainte temporelle diminue la probabilité d'interpréter une phrase sous-informative de façon pragmatique.

D'autres méthodes que la limitation du temps de réponse ont été utilisées pour manipuler expérimentalement l'effort cognitif des participants. De Neys & Schaeken (2007) ont par exemple proposé de limiter les ressources cognitives des participants en leur demandant d'effectuer une autre tâche cognitive en parallèle. Ils ont ainsi combiné une tâche de vérification de phrase classique, avec une tâche de mémoire de travail visuo-spatiale dans laquelle il était demandé aux participants de mémoriser une séquence de points dans une grille de 3 x 3 cases. La manipulation cruciale était que cette combinaison de 3 à 4 points pouvait être simple (3 points alignés) ou complexe (voir figure 1-3), mobilisant ainsi plus ou moins les ressources exécutives du participant. Les phrases à vérifier étaient insérées dans les items de la tâche de mémoire de travail, de sorte que pour chaque item une combinaison de points était présentée brièvement, puis la phrase que le participant devait évaluer comme correcte ou non était affichée, puis une grille vierge dans laquelle le participant devait rapporter la configuration de points initialement présentée.



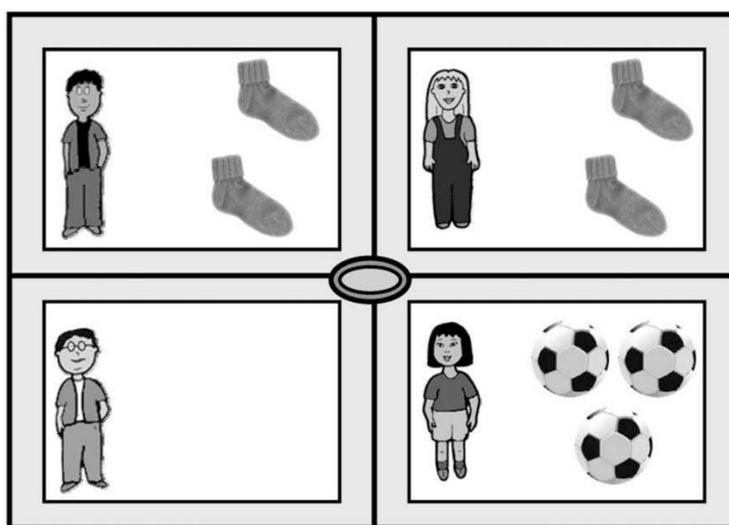
**Figure 1-3 :** Exemple de configuration de points complexe (A) ou simple (B) à mémoriser pendant la tâche d'implicature scalaires (illustration créée d'après De Neys & Schaeken, 2007)

Tandis que la charge cognitive n'avait pas d'incidence sur les jugements des phrases contrôles (corrects dans plus de 93% des cas), les réponses pragmatiques étaient significativement moins fréquentes lorsque la charge cognitive était élevée (73 % des réponses) que lorsqu'elle était faible (79 % des réponses). Ainsi, limiter les ressources cognitives des participants diminuait bien leur capacité à interpréter un terme scalaire de façon pragmatique, confirmant à nouveau l'effort associé à cette interprétation par rapport à une lecture littérale d'un tel énoncé, même si cet effet n'a pas toujours pu être répliqué (voir Fairchild & Papafragou, 2021).

Par ailleurs, Breheny et al. (2006) ont étudié le temps nécessaire au traitement d'expressions scalaires en elles-mêmes, au moment où elles étaient présentées (plutôt que le temps requis pour l'évaluation ultérieure de ces énoncés), et la mesure dans laquelle le contexte pouvait moduler leur interprétation. En effet, le contexte peut rendre plus ou moins pertinent l'exclusion d'une alternative plus informative. Par exemple, l'énoncé « J'ai invité *quelques* amis » peut être formulé en réponse à la question « As-tu invité tous tes amis ? » ou « As-tu prévu quelque chose pour ton anniversaire ? » ; dans le premier cas, la question est de savoir si tous les amis sont invités ou non, et l'accent est donc mis sur l'alternative plus informative 'tous', tandis que dans l'autre cas, la question est plutôt de savoir si des gens sont invités ou non, et l'accent est donc mis sur l'alternative moins informative 'aucun' (c'est la notion de *question en discussion*, Roberts, 1996). En exploitant ce contraste, Breheny et al. (2006) constituent des vignettes verbales se terminant par une expression scalaire et dont le contexte met l'accent ou bien sur l'alternative plus informative (favorisant l'implicature scalaire), ou bien sur l'alternative moins informative (décourageant l'implicature scalaire). En implémentant ces histoires dans un paradigme où les participants lisent ces histoires sur ordinateur, à leur propre rythme, ils peuvent mesurer le temps nécessaire pour lire une même expression avec un terme scalaire, en comparant les cas où le contexte encourage ou non l'implicature. Ils observent ainsi que les temps de lecture de ces expressions sont plus importants lorsque le contexte met l'accent sur l'alternative plus informative que dans l'autre cas (expérience 1 et 3). Cela confirme donc d'une part que le contexte module la dérivation des implicatures (voir plus bas l'étude de Zhao et al. [2021] pour une autre manipulation du contexte), et d'autre part que cette implicature est bien réalisée 'online' dès la présentation de l'expression, au prix d'un effort cognitif qui se traduit par un temps de lecture supérieur.

L'eye-tracking (ou oculométrie) permet également d'avoir des indices fins sur le déroulé temporel de ces processus. Huang et Snedeker (2009a) ont par exemple développé un paradigme de type 'visual-world' sensible aux implicatures scalaires. Dans cette tâche, des consignes sont présentées oralement à des participants tandis qu'ils explorent un environnement visuel et leurs mouvements oculaires sont enregistrés, permettant d'établir la direction de leur regard. Dans chaque item, 4 personnages (deux filles et deux garçons) se répartissent 2 types d'objets (par exemple des chaussettes et des ballons de foot, voir figure 1-4), un personnage possédant

l'ensemble des objets (par exemple, 3/3 ballons de foot) tandis que l'autre n'a qu'une partie des objets (par exemple, 2/4 chaussettes). Dans la condition critique, les participants entendent des consignes telles que « *Point to the girl who has some of the socks* » (montre la fille qui a quelques chaussettes). Cet énoncé est construit de façon à être progressivement désambiguïté : le mot 'girl' permet d'éliminer 2 des 4 situations. Puis, le mot 'some' (quelques) est sémantiquement compatible avec les 2 situations restantes, mais une interprétation pragmatique (quelques-mais-pas-tous) permet déjà de déterminer la cible. L'ambiguïté n'est définitivement levée qu'à la fin du mot 'socks' (chaussettes), puisque chaque mot et son distracteur ont la même syllabe initiale ('socks' et 'socket ball'), afin de prolonger autant que possible l'ambiguïté. Cette condition critique est comparée à des consignes telles que « *Point to the girl who has all of the socket-balls* » (montre la fille qui a tous les ballons de foot), dans lesquels le quantifieur 'all' n'est sémantiquement compatible qu'avec un seul personnage. La comparaison de ces deux situations permet donc de déterminer si l'interprétation pragmatique de 'quelques' est aussi rapide que les seuls processus sémantiques qui sont nécessaires dans la deuxième condition. D'autres conditions contrôle incluant des nombres exacts ne seront pas ici détaillées.



**Figure 1-4 :** Exemple d'item utilisé par Huang et Snedecker (2009a), dans lequel une instruction critique serait « *Point to the girl who has some of the socks* », comparé à « *Point to the girl who has all of the socket balls* »

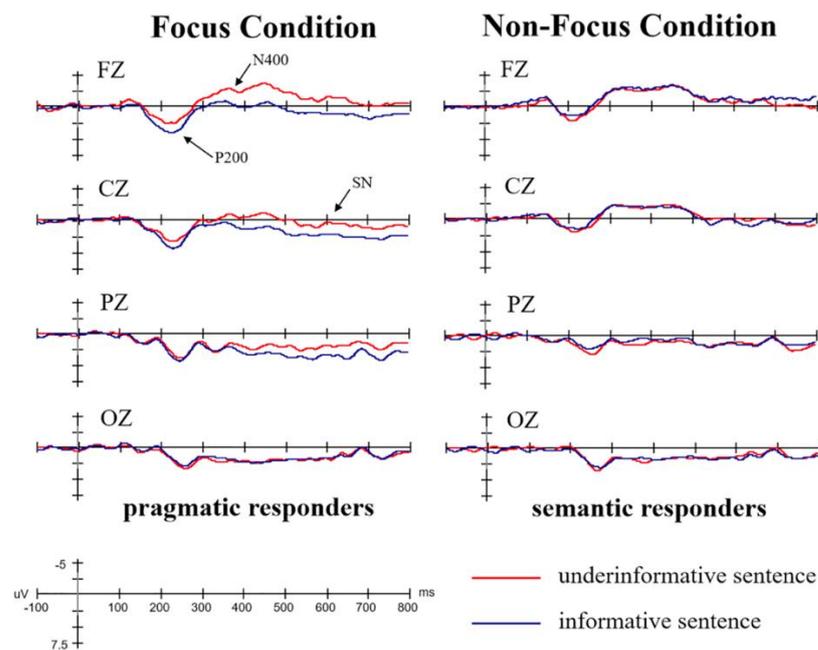
Alors que la quasi-totalité des réponses finales des participants sont correctes, les enregistrements de leurs mouvements oculaires confirment que dans les conditions contrôles, ils portent leur attention sur la cible dès la présentation du quantifieur (ils regardent donc préférentiellement la cible avant la fin de la phrase, dès 400 ms après la présentation du quantifieur). En revanche, dans la condition critique, les participants ne regardaient préférentiellement la cible que plus tardivement (à partir de 800 ms après le quantifieur). Ce résultat (confirmé dans deux expériences supplémentaires) indique donc que la désambiguïté est rapide lorsqu'elle se base sur des indices sémantiques, mais plus lente lorsqu'un enrichissement pragmatique est nécessaire,

suggérant que lors de la présentation d'un terme scalaire tel que 'quelques', l'accès au contenu sémantique précède l'enrichissement pragmatique.

Dans la gamme des techniques expérimentales qui ont été déployées pour évaluer l'effort cognitif associé à la dérivation d'implicatures scalaires, citons enfin les techniques neurophysiologiques, et en particulier les potentiels évoqués en électroencéphalographie (EEG), dont la haute résolution temporelle renseigne sur le déroulé temporel précis de cette opération. Dans une étude récente, Zhao et al. (2021) ont par exemple utilisé un paradigme de vérification de phrases (écrites). Plutôt que de demander aux participants d'évaluer si ces phrases étaient correctes sur la base de connaissances encyclopédiques, les auteurs ont proposé un support visuel sur la base duquel les phrases devaient être évaluées (un ensemble de carrés de différentes couleurs, contenant ou non des lettres, pour évaluer des phrases telles que « tous les carrés sont bleus ». Des phrases sous-informatives (e.g., « quelques carrés sont jaunes » quand tous les carrés sont jaunes) peuvent ainsi être classiquement comparées à des phrases correctes ou incorrectes. En plus d'enregistrer les potentiels évoqués par ce type de stimulus, cette expérimentation manipulait également le focus de la phrase sous-informative, en modifiant la nature des phrases contrôles erronées. Par exemple, à partir d'une situation de 6 carrés, dont 4 sont jaunes, et 3 ont une lettre, la phrase « 4 carrés ont une lettre » constitue une erreur sur la quantité, et met donc l'accent sur ce type d'information, tandis que la phrase « 4 carrés sont rouges » constitue une erreur sur la couleur, et ne met donc pas l'accent sur la quantité. Ainsi, l'attention des participants peut être implicitement guidée vers deux types d'informations lorsque des phrases sous-informatives telles que « quelques carrés sont jaunes » leur sont présentées : ou bien la quantité et donc le terme scalaires 'quelques', ou bien la couleur. Conformément à leurs hypothèses, les auteurs observent que lorsque le focus est placé sur le quantifieur, les termes scalaires sont interprétés de façon nettement plus pragmatique que lorsque ce n'est pas le cas (66% contre 22% de réponses pragmatiques), tandis qu'aucun effet n'est observé dans les conditions contrôles, très réussies.

Comme dans de nombreuses études (e.g. Bott & Noveck, 2004; Guasti et al., 2005; Noveck, 2001), les participants étaient le plus souvent cohérents dans leur façon de répondre, de sorte que les taux de réponses pragmatiques individuels étaient distribués de façon bimodale, avec un pic représentant les participants répondant toujours de façon pragmatique, et un autre pic de participants répondant toujours de façon sémantique. Lorsque le focus était mis sur le quantifieur,  $\frac{3}{4}$  de l'échantillon était constitué de répondants pragmatiques, et  $\frac{1}{4}$  de répondants sémantiques, avec peu de profils mixtes. Ces proportions étaient inversées dans la condition non-focalisée sur le quantifieur. Ainsi, placer le focus attentionnel sur le quantifieur incite probablement les participants à consacrer plus d'énergie à traiter ce terme, et ainsi à réaliser plus d'inférences pragmatiques.

Les potentiels évoqués par ces stimuli (figure 1-5) renseignent également sur leur traitement cérébral et cognitif. Dans tous les cas, le quantifieur ‘quelques’ (qu’il soit utilisé de façon informative ou non) a évoqué un composant P200, typiquement associé à l’allocation de ressources attentionnelles, et un composant N400, typiquement associé à un effort cognitif. Mais ces composants étaient plus amples chez les répondeurs pragmatiques dans la condition focalisée que chez les répondeurs sémantiques dans la condition non-focalisée. De plus, chez les répondeurs pragmatiques seulement, une négativité soutenue et tardive était observée en réponse aux phrases sous-informatives, et pas après l’utilisation appropriée de ‘quelques’, ce qui pourrait refléter les processus inférentiels complexes qui interviennent dans des situations d’ambiguïté. Dans l’ensemble, ces observations indiquent donc de façon convergente que les inférences pragmatiques de termes scalaires sont des opérations cognitivement coûteuses, modulées par le contexte et mise en œuvre dès la présentation du terme en question. D’autres potentiels évoqués apparaissaient également à la présentation du dernier mot, désambiguïsant la phrase, et indiquant que les processus inférentiels n’étaient pas uniquement déterminés par le terme scalaire lui-même.



**Figure 1-5 :** Potentiels évoqués par le quantifieur ‘quelques’ dans l’étude de Zhao et al. (2021), pour les répondeurs pragmatiques dans la condition ‘focus’ (à gauche) où sont visibles des composants P200, N400 et la négativité soutenue (SN) plus marquées en réponses aux quantifieur sous-informatifs (rouge) qu’informatifs (bleu), et les répondeurs sémantiques pour la condition ‘non-focus’ (à droite), où de telles différences sont absentes (figure extraite de la publication originale)

Cette courte revue illustre donc différentes approches expérimentales qui révèlent de façon concordante que l’enrichissement pragmatique de termes scalaires est généralement coûteux cognitivement, à l’inverse de ce que proposent les approches néo-gricéennes *par défaut*. L’effort cognitif et la propension des individus à réaliser des implicatures scalaires ne sont toutefois pas uniformes et sont notamment modulées

par le contexte, comme entrevu avec les études de Breheny et al. (2006) ou de Zhao et al. (2021), mais également par des facteurs syntaxiques tels que l'insertion dans des propositions conditionnelles (e.g. Hartshorne et al., 2015), par l'accessibilité des alternatives aux termes scalaires (e.g. Degen & Tanenhaus, 2015) ou par des variations interindividuelles comme les capacités de communication en général (e.g. Nieuwland et al., 2010). Ces études se focalisent toutefois sur le cas particulier des implicatures scalaires, ou éventuellement de phénomènes liés, mais ne prétendent pas s'appliquer aux inférences pragmatiques de façon générale. Avant d'examiner une source présumée de cet effort cognitif en [partie 3](#), nous proposons donc de nous tourner vers une littérature relativement distincte qui porte sur un autre type d'inférences pragmatiques : les métaphores.

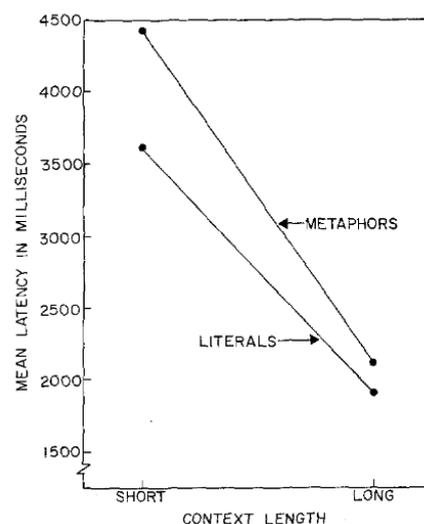
## 2.2. Métaphores

Les métaphores sont une illustration parmi les plus frappantes de disjonction entre le sens d'un énoncé et la signification que lui attribue son locuteur. Typiquement, il est impossible de ne s'appuyer que sur le sens littéral de l'énoncé « *Alex est une bombe à retardement* » pour comprendre l'intention de communication du locuteur. Il est dans ce cas nécessaire de sélectionner les propriétés pertinentes du *véhicule* (la bombe à retardement et le danger imminent que cela représente) et les transférer au *sujet* de cette expression (Alex et son humeur instable). Ce processus caractérise les métaphores (du grec *metaphorá*, littéralement : transport). Les métaphores se distinguent toutefois des comparaisons d'une part, et d'autres formes de langage figuratif comme l'hyperbole ou la métonymie d'autre part. En effet, dans les comparaisons le transfert entre deux concepts différents est explicité verbalement par un connecteur (e.g. « Alex est *comme* une bombe à retardement »), ce qui rend l'expression littéralement vraie, alors qu'une métaphore est littéralement fausse. Par ailleurs, les deux concepts qui sont associés dans une métaphore appartiennent à deux domaines différents (souvent, comme dans l'exemple de la bombe à retardement, un domaine concret est utilisé comme véhicule d'un domaine différent et plus abstrait) ; à l'inverse, la métonymie consiste à associer deux concepts appartenant au même domaine, liés par un lien conceptuel, en utilisant par exemple le contenant pour le contenu (e.g. « boire un *verre* »), le lieu pour l'institution qui y est installée (e.g. « *L'Élysée* est resté sourd aux manifestants »), un artiste pour désigner son œuvre (e.g. « Tu as vu le dernier *Spielberg* ? »), etc. L'hyperbole, enfin, s'appuie sur l'exagération d'un concept, dont l'interprétation nécessite *d'élargir* la signification littérale de l'expression à une catégorie plus large. Dans l'énoncé « Ça fait *mille ans* que je t'attends » par exemple, 'mille ans' doit être interprété comme l'expression plus globale d'une durée (trop) importante.

Parallèlement à l'étude des implicatures scalaires, une littérature importante s'est développée sur les processus permettant la compréhension des métaphores, et en particulier, suite aux travaux de Grice, sur la place qu'occupe le sens littéral de ces expressions. Nous évoquerons ici quelques-uns de ces travaux particulièrement

influent, qui mettent en évidence des facteurs et des mécanismes qu'on sait intervenir dans la compréhension des métaphores et qui sont mobilisés dans la littérature développementale (*chapitre 2*) ou pour l'étude de fonctionnements neurodivergents (*chapitre 3*).

Une façon d'évaluer la pertinence du Modèle Pragmatique Standard en 3 étapes est de comparer le traitement d'énoncés littéraux, mettant uniquement en œuvre la première étape du modèle (décodage linguistique de l'énoncé) à des énoncés métaphoriques, qui devraient être plus complexes puisqu'ils mettent en jeu des étapes supplémentaires permettant d'inférer les intentions de communication du locuteur. Cette approche a été adoptée dans les travaux princeps d'Ortony et al. (1978) ou de Gibbs (1978) qui ont utilisé des paradigmes de temps de lecture. Ortony et al. (1978) ont ainsi inséré dans une tâche de lecture des expressions dont l'interprétation pouvait être métaphorique ou littérale selon le contexte qui les précédaient. Par exemple, l'expression « *Malgré le danger, les troupes se mirent en marche* » était présentée après une histoire introduisant un contexte de guerre (favorisant une interprétation littérale) ou après une histoire mettant en scène un conflit entre des enfants et leur baby-sitter (induisant une interprétation métaphorique). Ces contextes pouvaient être complets ou minimaux (la première phrase seulement du contexte complet). Les histoires étaient présentées informatiquement, de sorte que les participants lisaient à leur rythme chaque expression. Cette procédure a révélé qu'une même expression était lue plus lentement lorsqu'un contexte minimal poussait à l'interpréter métaphoriquement que littéralement. Toutefois, un contexte plus riche facilitait et accélérail fortement la lecture de ces expressions, et la différence entre les interprétations métaphoriques et littérales n'était alors plus significative (figure 1-6).



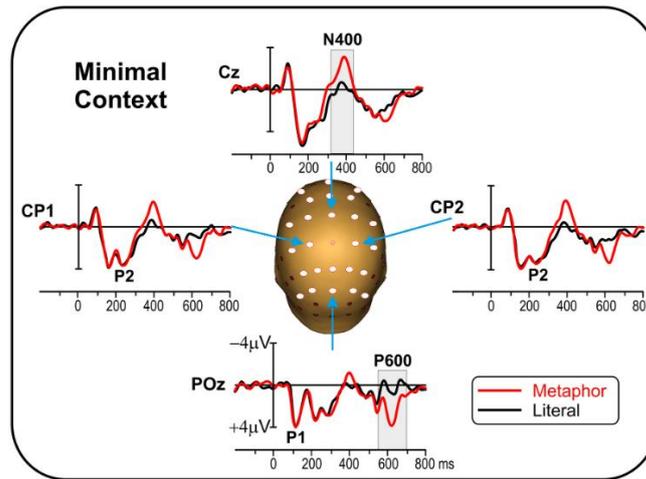
**Figure 1-6 :** Temps de lecture des participants de l'étude d'Ortony et al. (1978) pour les interprétations des expressions critiques, selon que le contexte induise une interprétation métaphorique ou littérale, et que le contexte soit pauvre ou riche (figure extraite de la publication originale)

Ces résultats démontrent donc qu'une métaphore peut dans certains cas être comprise aussi vite (et donc, facilement) qu'une expression littérale, ce qui va à l'encontre des

prédictions du SPM. Mais cela met aussi en évidence le rôle important du *contexte* dans la compréhension de ces expressions ; un contexte plus pauvre produit ainsi des résultats qui sont plus compatibles avec le SPM. Cette étude a ainsi été utilisée, parmi d'autres, pour soutenir un modèle alternatif dit d'*accès direct* qui postule que la signification métaphorique n'est pas cognitivement plus coûteuse que la signification littérale d'un énoncé et qu'elle peut ainsi être inférée *directement* tant qu'elle est soutenue par le contexte (Gibbs, 1994). Pour soutenir sa proposition, Gibbs s'appuie notamment sur une tâche légèrement différente (Gibbs, 1990), dans laquelle un référent est introduit dans une première partie de l'histoire, puis rappelé par une expression référentielle à la fin de l'histoire. Ces références pouvaient être métaphoriques ou synonymiques, et étaient également comparées à l'évocation de nouveaux référents. Si la lecture de la référence métaphorique était plus lente que celle de la référence synonymique, ces deux conditions ne différaient toutefois pas statistiquement de la condition contrôle. Ce paradigme s'appuie toutefois également sur un contexte riche qui, nous l'avons vu, semble moduler fortement les mécanismes de compréhension des métaphores.

Ce facteur 'contexte' a fait l'objet d'expérimentations plus récentes, par exemple par Bambini et al. (2016) qui ont utilisé l'électro-encéphalographie pour étudier les rôles respectifs du contenu littéral des expressions métaphoriques et du contexte. Dans une première expérience, des expressions littérales (e.g. évoquant un requin pour parler d'un poisson) et métaphoriques (e.g. évoquant un requin pour parler d'un avocat) ont été insérées dans des contextes minimaux et présentées à l'écrit à des participants dont l'électro-encéphalogramme était enregistré. La comparaison des potentiels évoqués par chaque type d'énoncé a révélé que, par rapport aux énoncés littéraux, les métaphores évoquaient des composants N400 et P600 de plus grande amplitude (figure 1.7), comme classiquement observé dans les études portant sur des métaphores isolées de type *X est Y* (e.g., certains avocats sont des requins ; pour une revue voir Canal & Bambini, 2023). Ces composants traduisent ainsi l'effort cognitif associé à la compréhension des métaphores, plus important que pour comprendre des expressions littérales équivalentes.

Dans une deuxième expérience les auteurs de l'étude ont présenté à des participants les mêmes stimuli insérés dans des contextes plus riches de deux phrases qui favorisaient l'interprétation des expressions cibles. Par rapport aux résultats de la première expérience, cette manipulation supprimait la différence observée entre les conditions au niveau du composant N400, mais pas du composant P600. Selon les auteurs, ces résultats éclairent donc comment le contexte faciliterait la compréhension des métaphores, en limitant l'effort d'accès aux représentations lexicales pertinentes, qu'indexerait le composant N400, sans toutefois supprimer l'effort nécessaire pour reconstruire l'intention de communication sous-jacente, qui serait indexé par le composant P600. Ce composant témoigne donc d'un effort cognitif nécessaire à la compréhension des métaphores, plus que pour une compréhension littérale, même avec un contexte riche et facilitateur.



**Figure 1-7:** Potentiels évoqués par des stimuli métaphoriques vs littéraux dans un contexte minimal dans l'étude de Bambini et al. (2016, expérience 1)

Le caractère conventionnel des métaphores est un autre facteur susceptible de moduler les mécanismes qui permettent de les comprendre. Si l'étude précédente employait seulement des métaphores non-conventionnelles, dites nouvelles, la répétition d'une métaphore dans l'utilisation de la langue peut conduire à sa lexicalisation, ce qui est susceptible de moduler son traitement. Bowdle et Gentner (2005) utilisent l'expression de « carrières des métaphores » pour décrire ce processus. Dans cette approche, une métaphore est initialement dite nouvelle, tant que l'association qu'elle établit entre le véhicule et le sujet de la métaphore est inédite (e.g. 'Cette femme est un pot de miel') ; puis sa répétition dans l'usage de la langue va conventionnaliser cette association (e.g. 'Cette femme est une sainte'). Cette métaphore peut continuer d'évoluer vers une *métaphore morte*, ou un idiom, dont le sens ne peut être interprété *que* de façon conventionnelle, et non plus en raisonnant à partir de ses éléments (e.g. 'Cette femme a le cœur sur la main'). Les mécanismes de traitements de ces différents types de métaphores seraient alors différents. Bowdle et Gentner rapprochent ainsi le modèle pragmatique standard et les tenants d'un accès direct, en postulant qu'il serait possible d'accéder *directement* au sens d'une métaphore conventionnelle, récupéré en mémoire, tandis qu'une métaphore nouvelle nécessiterait un accès indirect. En cohérence avec cette proposition, il a par exemple été montré que des métaphores nouvelles évoquaient des composants N400 plus amples que des métaphores conventionnelles (Arzouan et al., 2007; Lai et al., 2009), suggérant un effort cognitif plus important, et donc un accès moins direct à leur signification.

Une autre approche visant à éprouver le modèle pragmatique standard a consisté à examiner si la signification métaphorique d'une expression pouvait être interprétée automatiquement, plutôt qu'au prix d'un effort et secondairement au décodage de son sens littéral. Pour répondre à cette question, Glucksberg et al. (1982) ont exploité la propriété des métaphores d'avoir une signification pouvant être claire, tout en étant littéralement fausses. Ils ont présenté des métaphores de types '*certaines X sont Y*', telles que « *Certains boulots sont des prisons* » ou « *Certaines routes sont des*

*serpents* » à des participants, en leur demandant de juger si ces phrases étaient littéralement vraies ou non. Ces métaphores étaient comparées à des phrases littéralement vraies (e.g. « *Certains fruits sont des pommes* »), littéralement fausses mais non métaphoriques (e.g. « *Certains oiseaux sont des pommes* »), ainsi que d'autres phrases contrôles combinant les termes de plusieurs métaphores (e.g. « *certaines routes sont des prisons* »<sup>6</sup>), littéralement fausses également. Si les participants traitent d'abord le sens littéral d'une expression, avant de rechercher une signification métaphorique en cas de phrase littéralement fausse (modèle pragmatique standard), alors une métaphore et une phrase contrôle fautive devraient être rejetées avec autant de facilité. À l'inverse, si les participants accèdent directement et automatiquement à la signification métaphorique des expressions, alors cette signification devrait créer une interférence dans le jugement littéral de l'expression, qui devrait entraver le rejet des métaphores, par rapport aux phrases contrôles. Conformément à la deuxième hypothèse, Glucksberg et al. (1982) observent que les métaphores sont rejetées significativement plus lentement que les phrases contrôles, démontrant un *effet interférent des métaphores*. Ces observations suggèrent que l'accès au sens métaphorique d'une expression est automatique et irrépessible. Cet effet, initialement observé à partir de matériel écrit, a plus récemment été (partiellement) répliqué avec une présentation orale (Chouinard et al., 2019). À l'encontre du modèle pragmatique standard, ces travaux suggèrent donc un traitement cognitif des métaphores en au moins deux étapes, avec dans un premier temps la génération simultanée de significations littérales et métaphoriques, puis dans un second temps la sélection de l'interprétation pertinente, impliquant la suppression de la signification non-pertinente (voir aussi la figure 3-5 au [chapitre 3](#)).

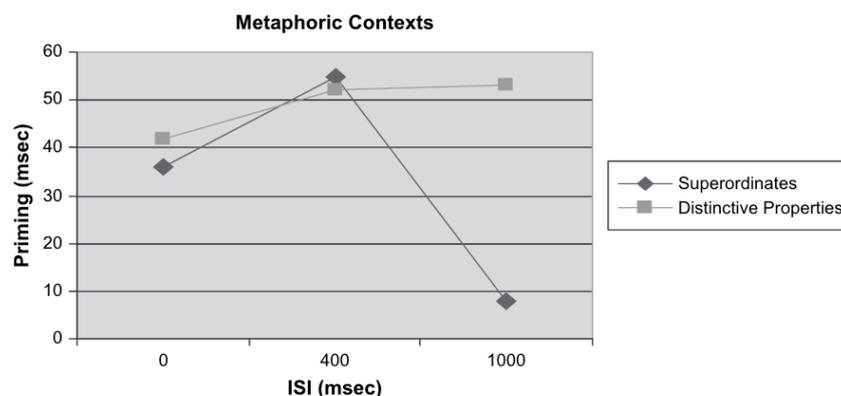
Cette proposition dépasse ainsi également l'opposition entre le modèle pragmatique standard et l'hypothèse d'un accès direct, en avançant qu'une interprétation métaphorique peut être générée de façon irrépessible, sans pour autant affirmer que le sens littéral de l'expression soit ignoré. Certaines études ont d'ailleurs montré que faciliter l'accès au sens littéral de l'énoncé, via un indice masqué, rendait le traitement d'une métaphore moins coûteux, tel qu'évalué en EEG par des potentiels évoqués (Weiland et al., 2014).

Rubio-Fernandez (2007) a plus directement évalué l'activation des différentes représentations associées au sens littéral d'un véhicule d'une métaphore, en proposant une tâche de décision lexicale (écrite) après la présentation de métaphores (orales) dans de brefs contextes. Par exemple, le mot 'cactus' était utilisé pour désigner une personne rebutée par les contacts physiques (« *John n'aime pas les contacts physiques. Même sa petite amie a des difficultés à se rapprocher de lui. John est un cactus* »). Ces métaphores étaient utilisées comme amorçage pour la tâche de décision lexicale, qui

---

<sup>6</sup> Exemple de phrase contrôle obtenue en combinant les métaphores « *certaines bouloles sont des prisons* » et « *certaines routes sont des serpents* ». Tous les exemples sont des traductions personnelles de l'étude originale en anglais.

incluait des mots liés aux propriétés génériques du véhicule, non-pertinentes dans le contexte métaphorique (dans cet exemple ‘plante’), ou des mots liés aux propriétés du véhicule pertinentes dans ce contexte (ici, ‘piquant’). Des mots sans relation avec le véhicule, utilisés dans d’autres métaphores étaient ajoutés comme contrôle (par exemple pour la métaphore du cactus, ‘coûteux’). Si une représentation est activée par la présentation de la métaphore, son traitement devrait être facilité (et donc plus rapide) dans la tâche de décision lexicale, par rapport à un mot sans aucun rapport. Rubio-Fernandez (2007) montre ainsi qu’immédiatement après la présentation d’une métaphore, le traitement des mots renvoyant aux différentes propriétés sémantiques du véhicule est facilité, par rapport aux mots sans relation, et ce autant pour les propriétés pertinentes que non-pertinentes (e.g. la métaphore du cactus facilite autant le traitement de ‘plante’ que de ‘piquant’). L’ensemble des représentations sémantiques littéralement associées au véhicule sont donc activées lors de la présentation d’une métaphore. Toutefois, Rubio-Fernandez manipule également le délai entre la présentation de la métaphore et la décision lexicale. Ce faisant, elle démontre que si l’effet observé immédiatement après la présentation de la métaphore est maintenu 400 ms après la métaphore, ce n’est plus le cas après un délai de 1000 ms (voir figure 1-8). En effet, une seconde après la présentation de la métaphore, le traitement de la propriété pertinente du véhicule (‘piquant’) reste facilité par rapport à un mot sans rapport, mais plus le traitement des propriétés génériques non-pertinentes (‘plante’). Pour Rubio-Fernandez (2007), ces observations indiquent qu’une métaphore active initialement l’ensemble des propriétés littérales associées au véhicule, puis que dans un second temps les propriétés non-pertinentes sont « *activement supprimées* » en fonction du contexte.



**Figure 1-8:** Effet facilitateur de la métaphore sur le traitement de mots liés aux propriété pertinentes (*distinctive properties*) ou non-pertinentes (*superordinates*) par rapport aux mots non-liés, à différents délais après la présentation de la métaphore (ISI) (figure extraite de Rubio-Fernandez, 2007)

En somme, ces exemples de travaux illustrent donc comment la vision pragmatique standard a pu être enrichie, en montrant que si l’accès à la signification métaphorique d’une expression peut être irrésistible, il demeure plus coûteux que l’accès aux représentations littérales, demandant notamment un effort de suppression des propriétés littérales non-pertinentes. Ces processus sont par ailleurs modulés par des

facteurs contextuels, ou propres à l'expression métaphorique, comme sa conventionalité.

D'autres travaux et approches théoriques se sont focalisées sur les processus spécifiquement mobilisés dans la compréhension des métaphores (voir e.g. la revue de Holyoak & Stamenković, 2018). Certaines approches proposent ainsi que la compréhension des métaphores s'appuie sur un processus de *catégorisation*, le véhicule étant interprété d'une façon plus large que son sens littéral (sous la forme d'une catégorie), sous la contrainte du sujet de la métaphore (Glucksberg & Keysar, 1990) ; d'autres évoquent plutôt un processus d'*analogie* entre les deux concepts connectés (Gentner & Clement, 1988; Tourangeau & Sternberg, 1982). D'autres approches encore proposent que les métaphores dans le langage soient ancrées sur un système métaphorique conceptuel qui structure plus largement la pensée humaine (et pas seulement le langage), permettant de manipuler des objets abstraits sous la forme d'images concrètes (Lakoff & Johnson, 1980). Cette thèse portant sur les mécanismes plus généraux des inférences pragmatique, nous nous concentrerons toutefois sur des travaux moins spécifiques que ceux-ci.

En nous focalisant dans cette partie sur la place des représentations littérales dans le traitement des implicatures scalaires comme des métaphores, nous avons vu comment le modèle pragmatique standard (issu des propositions gricéennes d'un principe de coopérations et de maximes conversationnelles) avait pu être nuancé et augmenté, par des approches expérimentales qui précisent l'effort cognitif associé aux inférences pragmatiques. Nous examinerons dans la partie suivante une source potentielle de cet effort, en évoquant d'autres approches théoriques et expérimentales issues des travaux de Grice.

### 3. Inférences pragmatiques et théorie de l'esprit

La proposition fondamentale de Grice que la communication humaine consiste à transmettre et reconnaître des *intentions* a fait l'objet de développements théoriques particuliers. Cette base cognitive de la communication, le raisonnement sur les états mentaux d'autrui, désignée par l'expression de *théorie de l'esprit*<sup>7</sup>, est elle-même l'objet d'une littérature importante en psychologie (voir [encart 1](#)). Dans cette partie, nous développerons la proposition théorique de la *théorie de la pertinence* puis évoquerons des données expérimentales qui soutiennent le lien entre pragmatique et théorie de l'esprit, avant d'apporter quelques nuances théoriques qui complexifient cette relation.

---

<sup>7</sup> Plusieurs expressions plus ou moins équivalentes coexistent en anglais : *theory of mind*, *mindreading*, *mentalizing*, *folk psychology* ; cette multiplicité est à notre connaissance bien moindre en français, aussi utiliserons nous le terme de *théorie de l'esprit* comme unique traduction de ces concepts.

### **Encart 1** : La théorie de l'esprit, conceptualisation et opérationnalisation

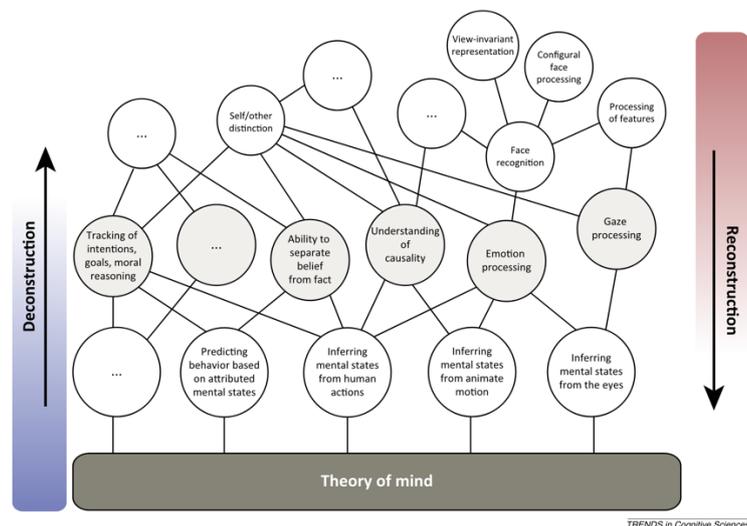
Le concept de *théorie de l'esprit* (Premack & Woodruff, 1978), issu de travaux en éthologie, fait référence à la capacité à attribuer à soi-même et à autrui des états mentaux tels que des intentions, des savoirs, des croyances, et à utiliser ces représentations pour comprendre et prédire leurs comportements. Cette compétence est essentielle pour les êtres humains, constituant la base de la représentation de soi, de la coopération et du raisonnement (Waytz et al., 2010). D'un point de vue expérimental, ce concept est typiquement opérationnalisé comme la capacité à comprendre de *fausses croyances*. Non pas que la théorie de l'esprit ne soit pas nécessaire pour comprendre de vraies croyances, mais la distinction entre le monde et la représentation du monde qu'un individu détient serait alors masquée. En effet, un individu se comportera différemment s'il se base sur une représentation erronée du monde plutôt que sur le monde tel que nous le connaissons, tandis qu'une croyance correcte donnerait lieu au même comportement. Ainsi, la capacité à se représenter explicitement les croyances d'autrui est nécessaire pour pouvoir comprendre que la croyance d'un individu est fautive (Dennett, 1978).

C'est sur la base de ce raisonnement que Wimmer et Perner (1983) ont développé le paradigme expérimental classique permettant d'évaluer la théorie de l'esprit. Ils ont proposé à des enfants des situations dans lesquelles un personnage, Maxi, range un chocolat dans un placard X. En son absence, sa mère le déplace dans un placard Y. Maxi croit donc (à tort) que le chocolat se trouve toujours dans le placard X, alors que les participants savent qu'il se trouve dans le placard Y. Demander aux participants où ce personnage irait chercher le chocolat permet ainsi d'établir s'ils sont capables de se représenter les croyances du personnage. Dans cette étude princeps, aucun des enfants de 3-4 ans n'a pu répondre correctement à cette question, tandis que 57 % des enfants de 5-6 ans répondaient correctement, le taux de réussite atteignant 86% à 6-9 ans. Cette étape importante dans la compréhension des fausses croyances aux alentours de 4-5 ans a été confirmée par de nombreuses études (Wellman et al., 2001). Cette tâche a donné lieu à une littérature importante en psychologie du développement, qui sera évoquée dans un [encart du chapitre 2](#).

Depuis son introduction, le concept de théorie de l'esprit s'est toutefois complexifié et n'est plus considéré de façon monolithique, mais plutôt comme un ensemble de processus (voir Happé et al., 2017; Schaafsma et al., 2015), au sein desquels des distinctions importantes ont été postulées. On distingue notamment des aspects plutôt élémentaires ou basiques comme l'attribution d'intentions de premier ordre (telles qu'évaluée dans le paradigme de Maxi évoqué ci-dessus) et des aspects plus complexes comme la représentation de croyances de deuxième ordre (Pierre pense que Marie croit que...), ou éventuellement d'ordre supérieur encore. On oppose également des représentations plutôt cognitives (compréhension d'états mentaux tels que les croyances, les connaissances, les perspectives) à des représentations plus affectives (émotions, sentiments, attitudes), ou encore les représentations d'autrui et de soi, ou des processus automatiques et rapides à d'autres processus volontaires et coûteux (voir figure 1-9).

Si cette complexité n'est pas au centre de cette thèse, elle ne peut être ignorée car elle se traduit par une multiplicité de tâches expérimentales qui sont utilisées pour opérationnaliser la théorie de l'esprit (voir par exemple Beaudoin et al., 2020, pour les tâches utilisées avec les enfants de moins de 6 ans). Outre la tâche standard de fausses croyances évoquée ci-dessus, nous détaillerons ici à titre d'exemple deux tâches expérimentales qui seront utiles pour cette thèse. Premièrement, la tâche *Reading the Mind in the Eyes* (RME, Baron-Cohen et al., 2001) est conçue pour cibler les aspects plus

complexes et émotionnels de la théorie de l'esprit ; dans cette tâche, des photographies de la région des yeux de différentes personnes sont présentées aux participants, qui doivent sélectionner parmi 4 propositions celle qui représente le mieux ses états mentaux. Ces propositions sont des expressions écrites, parfois complexes, telles que « joueur », « reconnaissant », « contemplatif » ou « bouche bée » (traductions francophones de Prevost et al., 2014). Une autre tâche très utilisée est celle des *Strange Stories* (Happé, 1994; O'Hare et al., 2009), qui évalue la théorie de l'esprit dite « avancée » à l'aide d'histoires impliquant différents personnages. Chaque histoire implique un personnage qui produit un énoncé littéralement faux, ce qui s'explique par ses intentions ou ses croyances ; il peut ainsi mentir (qu'il s'agisse de pieux mensonges ou non), se tromper, faire semblant, être sarcastiques, bluffer, etc. Après chaque histoire, deux questions sont posées au participant, une question fermée demandant si la déclaration du personnage est vraie ou non, et une question ouverte demandant au participant d'expliquer pourquoi le personnage a produit un tel énoncé. Cette tâche nécessite donc de mobiliser la compréhension d'états mentaux relativement complexes, pour interpréter et expliquer correctement le comportement de personnages (mais voir aussi les autres compétences que cela mobilise dans la [section 3.3](#)).



**Figure 1-9 :** Illustration des processus et comportements multiples qui sont regroupés sous le concept de théorie de l'esprit et que Schaafsma et al. (2015) appellent à déconstruire, puis à reconstruire (extrait de la publication originale)

### 3.1. La théorie de la pertinence

À la différence des approches évoquées en deuxième partie, la *théorie de la pertinence* de Dan Sperber et Deirdre Wilson (Sperber & D. Wilson, 1996; D. Wilson & Sperber, 2006) propose une description globale du fonctionnement cognitif qui, appliquée à la communication, fournit une procédure générale permettant d'inférer les intentions de communication d'un locuteur. Au cœur de cette proposition se trouve la notion de *pertinence* d'un stimulus, qui est définie comme maximisée quand :

- (i) les *effets* cognitifs positifs du traitement de ce stimulus sont importants, c'est-à-dire quand ce traitement modifie de façon *utile* pour un individu ses

représentations, dans une situation donnée, par exemple en répondant à une interrogation, un doute, un besoin ;

(ii) les *coûts* cognitifs que ce traitement nécessite sont faibles.

Ainsi, sur la base de ces deux facteurs (effets et coûts), la pertinence de plusieurs stimuli peut être comparée, et l'individu pourrait s'orienter vers ceux qu'il estime les plus pertinents (*principe cognitif de la pertinence*). Dans ce cadre général de la cognition, Sperber et Wilson proposent un *principe communicatif de pertinence*, postulant que tout stimulus communicatif ostensif véhicule une présomption de sa propre *pertinence optimale*. Ainsi, en termes d'effets et de coûts, le fait même de réaliser un acte ostensif suscite chez l'auditeur une attente que ce stimulus vaille l'effort d'être traité d'une part, et qu'il soit le plus pertinent qui soit d'autre part (dans la limite des compétences et préférences du locuteur). Par exemple, en donnant à lire ce chapitre de thèse (acte de communication ostensif), je communique et suscite la présomption qu'il sera optimalement pertinent pour transmettre mon intention de communication, dans les limites de ma capacité à construire un tel discours. Ainsi, j'ai estimé les coûts qu'induiront la lecture de ce chapitre et annonce en le donnant à lire qu'ils seront contrebalancés par des effets positifs qu'induiront ces informations, par exemple sur la compréhension des chapitres ultérieurs. Le lecteur, quant à lui, pourra utiliser cette présomption pour inférer aussi précisément que possible l'intention qui était probablement la mienne lorsque j'ai écrit ces signes dans mon manuscrit. Le principe communicatif de pertinence permet en effet d'établir une procédure générale qu'un auditeur utiliserait pour inférer les intentions de communication d'un locuteur. Cette procédure consiste à rechercher des effets positifs d'un énoncé en suivant la voie du moindre effort, en combinant différentes hypothèses et attentes qui peuvent être formées à partir de la situation, et à arrêter ce processus dès que les attentes de pertinence sont satisfaites.

Sans qu'il n'y ait plus besoin du principe de coopération de Grice ni de ses déclinaisons en maximes conversationnelles, cette procédure générale s'applique ainsi à tous les types d'implicatures (et même à la reconstruction du sens littéral d'un énoncé). Dans le cas d'une implicature scalaire, l'exemple d'un locuteur qui affirmerait « *J'ai mangé quelques gâteaux* » peut être interprété comme signifiant qu'il n'a *pas* mangé tous les gâteaux, car dans le cas contraire l'énoncé « *j'ai mangé tous les gâteaux* » aurait été plus *pertinent* (ici, informatif). Si l'énoncé produit est le plus pertinent, c'est alors que le locuteur n'a pas mangé tous les gâteaux. Dans le cas d'une métaphore, la théorie de la pertinence peut par exemple être mobilisée pour comprendre l'énoncé « *Écrire une thèse est un marathon auquel il vaut mieux être bien préparé* »<sup>8</sup>. Le mot 'marathon' active ici un ensemble de connaissances encyclopédiques autour du concept de marathon (ce en quoi cela consiste, qui y participe, quel type d'effort cela demande,

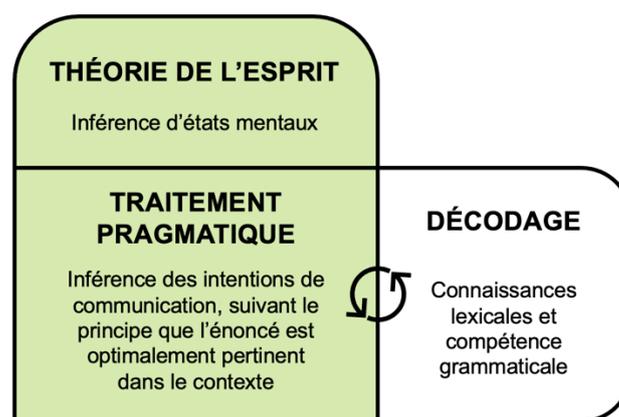
---

<sup>8</sup> Exemple adapté de Carston et Wearing (2011) ; pour un développement du parallèle entre une thèse et un marathon, voir le billet de blog de Jayron Habibe (2023)

etc.). Le contexte dans lequel ce terme est utilisé (l'écriture d'une thèse) et la connaissance du locuteur de ce contexte (Wearing, 2010) rend certaines de ces informations très accessibles (e.g. les notions d'effort, de durée, de détermination) tandis que d'autres le sont moins (la distance exacte d'un marathon, la présence de concurrents, etc.). Ainsi, interpréter cet énoncé par *la voie du moindre effort* consiste à utiliser ces informations les plus accessibles et à interpréter marathon comme signifiant quelque chose de long, nécessitant efforts et détermination, pouvant nécessiter une préparation importante, mentale comme matérielle. Cette interprétation étant suffisamment pertinente par rapport à ce qui est attendu, le processus inférentiel peut être stoppé.

On voit ainsi comment l'utilisation du langage, en expression comme en compréhension, mobiliserait des raisonnements métapsychologiques sur les représentations d'autrui ou, en d'autres termes, comment la pragmatique serait une manifestation de la théorie de l'esprit. Après avoir émis cette proposition initiale, les auteurs de la théorie de la pertinence ont plus tard adopté sur cette question une perspective modulaire, en proposant que la compréhension pragmatique soit prise en charge par un sous-module spécialisé de la théorie de l'esprit, qui opèrerait de façon spontanée et intuitive, plutôt que consciente et réflexive (D. Wilson & Sperber, 2006).

Cette théorie a eu un impact considérable sur les développements et expérimentations ultérieures. Elle a par exemple été reprise à leur compte par A.C. Wilson et Bishop (2019, voir figure 1-10) dans leur développement d'une mesure des inférences pragmatiques, démontrant à l'aide d'analyses factorielles de données obtenues chez 120 adultes neurotypiques que le décodage d'une part, et le traitement pragmatique associé à d'autres épreuves de compréhension sociale d'autre part, constituaient bien deux facteurs distincts, donc deux compétences distinctes, bien qu'en interaction pour traiter le langage en situation réelle.



**Figure 1-10 :** Modèle de l'interprétation du langage d'A.C. Wilson et Bishop (2019) reprenant la théorie de la pertinence (traduction personnelle)

L'influence théorique majeure du modèle de communication ostensif-inférentiel et de la théorie de la pertinence est aussi visible dans des approches computationnelles plus

récentes, telles que le modèle du *Rational Speech Act* (RSA, Degen, 2023; Goodman & Frank, 2016). Le RSA propose de modéliser informatiquement l'interprétation d'énoncés verbaux comme une opération bayésienne visant à déterminer l'intention de communication la plus probable d'un locuteur, étant donné ses productions verbales. Ce paradigme propose de modéliser un compreneur, qui modéliserait un locuteur lui-même, qui modéliserait à son tour comment son énoncé serait interprété par un compreneur littéral. Cette implémentation computationnelle permet ainsi d'intégrer dans un algorithme le sens littéral d'un énoncé d'une part (via le compreneur littéral), avec des inférences sur les états mentaux d'autrui d'autre part (notamment via les modélisations récursives), pour prédire comment des énoncés seraient interprétés. Cette approche a permis de modéliser l'interprétation d'une diversité d'énoncés pragmatiques, prédisant notamment avec succès comment des participants humains interprètent des implicatures scalaires (Goodman & Stuhlmüller, 2013) ou le langage figuratif (Kao et al., 2014), démontrant ainsi qu'implémenter un raisonnement sur les états mentaux d'autrui permet à une machine de distinguer ce qu'un locuteur *dit* de ce qu'il *veut dire* (mais voir aussi van Arkel et al. [2020] sur les coûts computationnels majeurs de ces modèles).

Compréhension pragmatique et théorie de l'esprit sont donc articulées par des liens forts au niveau théorique, ces liens étant par ailleurs soutenus par un ensemble d'observations empiriques, notamment chez l'adulte.

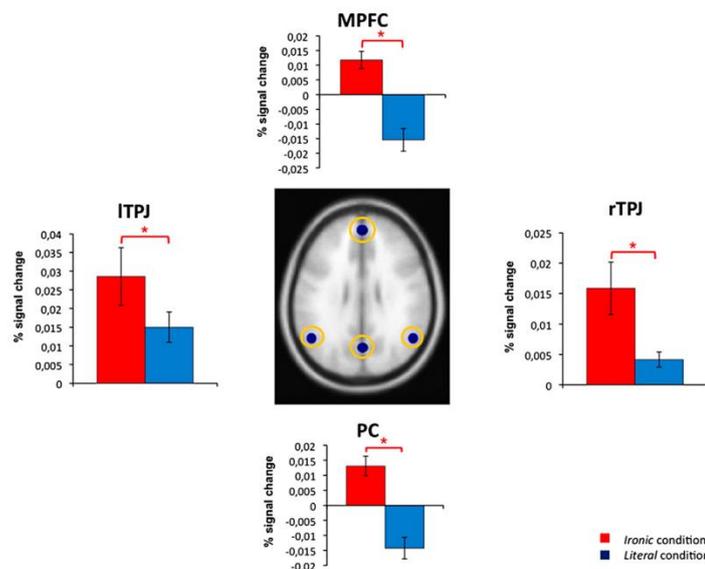
### **3.2. Pragmatique et théorie de l'esprit chez l'adulte**

Les liens entre pragmatique et théorie de l'esprit sont en effet soutenus par un ensemble de données, notamment chez l'adulte, que nous évoquerons ici avant d'évoquer celles portant sur le développement (*chapitre 2*) ou des populations atypiques comme l'autisme (*chapitre 3*), plutôt que de suivre une logique chronologique<sup>9</sup>. Ces liens ont notamment été investigués sous l'angle des neurosciences, en recherchant l'implication d'aires cérébrales typiquement associées à la théorie de l'esprit dans des tâches pragmatiques. Par exemple, Spotorno et al. (2012) ont évalué l'activité cérébrale associée au traitement d'énoncés ironiques vs non-ironique en IRM fonctionnelle. Pour ce faire, les auteurs ont construit des histoires écrites que les participants lisaient à leur rythme. À la fin de ces histoires, un personnage produisait un énoncé cible qui pouvait être interprété de différentes façons. Les histoires étaient en effet construites par paires, différant uniquement dans quelques détails du contexte, qui induisaient une interprétation littérale ou ironique du même énoncé cible à la fin de l'histoire. Ces histoires ont été mélangées à des histoires *fillers*, assurant que les items ironiques demeuraient peu fréquents et imprédictibles. Elles ont ensuite été proposées à des participants adultes pendant l'enregistrement de leur activité cérébrale

---

<sup>9</sup> Nous explorerons finalement ces liens à rebours de l'ordre dans lequel ils ont émergé dans la littérature, dans une certaine mesure, les travaux chez l'adulte s'appuyant souvent sur des travaux réalisés chez l'enfant, eux-mêmes souvent basés sur l'étude de fonctionnements atypiques.

en IRMf. Une question sur un détail de l’histoire était ensuite posée aux participants, afin de les engager à prêter attention à chaque histoire. Les résultats ont montré, d’abord au niveau comportemental, que les énoncés cibles étaient lus plus lentement lorsque le contexte poussait à les interpréter comme ironiques que littéralement, signant un effort particulier associé à cette condition. C’est la nature de ces efforts que les données neurofonctionnelles permettaient d’étudier. Celles-ci ont révélé que dans 4 régions cérébrales pré-identifiées comme fortement associées à la théorie de l’esprit (le cortex préfrontal médian, les jonctions temporo-pariétales gauche et droite et le précunéus), les énoncés cibles ironiques induisaient une activité plus importante que les énoncés contrôles littéraux (Figure 1-11).



**Figure 1-11** : Figure extraite de Spotorno et al. (2012) et représentant la suractivation cérébrale associée aux énoncés ironiques par rapport aux énoncés littéraux dans 4 régions cérébrales typiquement associées à la théorie de l’esprit (cortex préfrontal médian = MPFC, jonctions temporo-pariétales gauche et droite = ITPJ et rTPJ et précunéus = PC).

Par ailleurs, la condition ironique était associée à une plus grande connectivité fonctionnelle entre la partie ventrale du cortex préfrontal médian, dans le réseau de la théorie de l’esprit, et le gyrus frontal inférieur gauche, typiquement impliqué dans le traitement du langage. Ces données mettent donc en évidence l’implication du réseau cérébral de la théorie de l’esprit dans le traitement d’énoncés ironiques, par rapport à des énoncés littéraux, en interaction avec le réseau du langage. Une méta-analyse récente (Hauptman et al., 2023) rassemblant les données d’IRM fonctionnelle de 1 430 participants dans des tâches de compréhension pragmatique variées (rassemblant sans distinction ironie, métaphores, requêtes indirectes, métonymies, idiomes, proverbes, etc.), confirme l’implication des réseaux cérébraux du langage et de la théorie de l’esprit dans la compréhension pragmatique, par rapport au simple décodage d’énoncés littéraux. Le contraste inverse, entre les énoncés littéraux et pragmatiques, mobilisait moins les réseaux associés à l’effort cognitif en général et au contrôle exécutif, renforçant la spécificité de l’implication de la théorie de l’esprit dans la compréhension pragmatique.

À la suite de leur étude neurofonctionnelle de l'ironie, Spotorno et Noveck (2014) se sont focalisés sur les résultats comportementaux qu'ils avaient obtenus, et ont montré que le ralentissement associé à la condition ironique (par rapport aux mêmes énoncés dont le contexte induisait une interprétation littérale) pouvait être modulé. En effet, il diminuait progressivement lorsque les conditions expérimentales rendaient la survenue de l'ironie prédictible, et il était alors seulement observé au début de la tâche. Pour ces auteurs, cela montre que le ralentissement n'est pas seulement associé à l'énoncé ironique en lui-même, mais à l'effort de théorie de l'esprit nécessaire au début de la tâche, pour inférer les intentions de communication du locuteur, qui n'est plus nécessaire lorsque celui-ci est devenu prévisible, à la fin de la tâche. Ce résultat est notable car il a été transposé aux implicatures scalaires, dont les liens avec la théorie de l'esprit ont fait l'objet d'une littérature à ce jour plus pauvre.

Ronderos et Noveck (2023) ont ainsi reproduit l'effet de Bott & Noveck (2004) avec une version web de cette tâche, répliquant le ralentissement associé aux interprétations pragmatiques d'énoncés sous-informatifs, par rapport à une interprétation logique et à des énoncés contrôles (expérience 1). Puis ils ont démontré que, si les réponses pragmatiques étaient plus coûteuses au début de l'expérience, ce n'était plus le cas à la fin de l'expérience (expérience 2), révélant que les expressions sous-informatives ne nécessitent pas un effort particulier *en elles-mêmes*, mais que cet effort est lié à la situation. Ronderos et Noveck proposent alors, comme pour l'ironie, que ce ralentissement traduit l'effort nécessaire pour accéder aux intentions de communication du locuteur, et que celui-ci diminue dès lors que ce locuteur devient connu et donc prédictible. Pour tester cette proposition, ils modifient leur paradigme (expérience 3) de façon à ce que les premiers énoncés soient produits par un premier locuteur, puis qu'un second locuteur produise les derniers énoncés. Selon leur hypothèse, l'introduction d'un nouveau locuteur devrait rétablir un surcoût temporel associé aux réponses pragmatiques, puisque les intentions d'un nouveau locuteur sont par principe inconnues. À l'inverse, si la diminution du surcoût au fur et à mesure de la tâche était attribuable à un simple effet d'habituation, l'intervention d'un nouveau locuteur ne devrait pas avoir d'incidence sur cet effet. Leurs résultats sont bien en faveur de leur hypothèse, puisque l'introduction d'un nouveau locuteur au moment où les réponses pragmatiques étaient devenues aussi rapides que les réponses contrôles rétablissent un ralentissement, comme s'il devenait nécessaire de s'adapter aux intentions de communication d'une nouvelle personne.

Une autre approche consiste à s'intéresser aux différences interindividuelles dans les performances à des tâches pragmatiques et à des tâches de théorie de l'esprit. C'est celle qu'ont adoptée Fairchild et Papafragou (2021), en utilisant auprès de 178 participants une tâche de jugement binaire classique pour évaluer les implicatures scalaires, deux tâches de théorie de l'esprit (des extraits des tests Reading the Mind in the Eyes de Baron Cohen et al., 2001, et des Strange Stories de Happé, 1994), agrégées dans un score composite, et deux tâches de mémoire de travail (visuelle et verbale) évaluant les fonctions exécutives. Ils observent ainsi que les performances en théorie

de l'esprit sont bien corrélées aux implicatures scalaires ; les individus avec une meilleure théorie de l'esprit fournissent plus d'interprétations pragmatiques de termes scalaires que les autres, suggérant l'implication de la théorie de l'esprit dans la dérivation de cette inférence pragmatique. De plus, une fois ce facteur pris en compte, les fonctions exécutives ne contribuaient pas significativement aux variations dans la tâche d'implicatures scalaires. Ce résultat a été répliqué dans une seconde étude auprès de 174 nouveaux participants, et étendu à d'autres phénomènes pragmatiques. Les mêmes observations ont en effet été réalisées dans une tâche évaluant les requêtes indirectes. Cette seconde expérience incluait également une tâche de métaphore, dans laquelle il était demandé aux participants d'évaluer si des phrases (métaphoriques ou littérales) avaient du sens ou non. Cependant, les performances à cette tâche étaient bien corrélées aux scores de théorie de l'esprit, mais de façon contre-intuitive. En effet, les participants avec une meilleure théorie de l'esprit ont jugé les métaphores qui leur étaient présentées comme ayant *moins* de sens que ceux avec une moins bonne théorie de l'esprit. Pour comprendre ce résultat, il faut garder à l'esprit que cette tâche n'évaluait pas directement la compréhension de ces métaphores. Les auteurs avancent que leurs métaphores pouvaient être jugées comme ayant moins de sens que des phrases littérales, non pas parce qu'elles n'étaient pas comprises, mais parce que leur nature non-littérale était repérée. Ainsi, ils proposent que les participants avec une meilleure théorie de l'esprit seraient plus sensibles à la distinction entre significations littérales et non-littérales, et ainsi plus enclins à participer à cette tâche avec l'intention de repérer de telles distinctions. Il faut par ailleurs noter une limite importante de cette étude, qui a (en partie) évalué la théorie de l'esprit à l'aide d'extraits des *Strange Stories* (Happé, 1994), dont certains items se focalisent sur la compréhension de figures de style, de métaphores, de l'ironie, et qui mobilisent donc directement la pragmatique. Utiliser cette tâche pour étudier les liens entre théorie de l'esprit et pragmatique crée donc un risque de tautologie important, même s'il est ici limité par le fait que d'autres items étaient inclus, et que cette tâche était combinée avec le test RME. Nous aurons l'occasion de revenir sur cette limite méthodologique à la fin de ce chapitre.

Combinant une approche des différences interindividuelles avec une mesure neurophysiologique, Canal et collaborateurs (2022) ont étudié les potentiels évoqués par des énoncés métaphoriques. Dans cette étude, les EEG de 35 participants adultes ont été enregistrés durant la présentation orale de métaphores de type 'X est Y', dans lesquelles des êtres humains étaient associés à des entités non-humaines, ces associations pouvant reposer sur des caractères physiques (e.g. *I gemelli sono fotocopia*, Les jumeaux sont des photocopies) ou psychologiques (e.g. *I filosofi sono aeroplani*, Les philosophes sont des avions). Les participants étaient par ailleurs évalués avec deux tâches de théorie de l'esprit, le test RME (Baron-Cohen et al., 2001) et la tâche des triangles animés (Abell et al., 2000) évaluant l'attribution d'intention. Les résultats révélaient que les scores au test RME (et pas à la tâche des triangles) modulaient le composant N400 évoqué par les métaphores et qui trahit typiquement l'effort cognitif qu'elles induisent (Canal & Bambini, 2023). De meilleures

performances en théorie de l'esprit étaient associées à une N400 de moindre amplitude, indiquant que le traitement des métaphores induisait chez ces participants un effort moindre. Tandis que cet effet était observé de façon générale sur tous les types de métaphores, des subtilités apparaissaient également lorsque les métaphores physiques et psychologiques étaient contrastées dans les analyses exploratoires. Celles-ci ont révélé que l'effet de la théorie de l'esprit serait légèrement plus précoce et plus important pour les métaphores psychologiques que pour les métaphores physiques. Ces observations renforcent donc l'hypothèse que la théorie de l'esprit, au moins dans ses aspects plus complexes et émotionnels tels qu'évalués par le test RME, serait impliquée dans le traitement cérébral des métaphores, et que cette implication serait modulée par le type de métaphore en question, avec un engagement plus fort en cas de métaphore psychologique.

Certaines études ont par ailleurs directement mesuré l'influence des états mentaux d'un locuteur sur l'interprétation de ses énoncés par des participants en manipulant expérimentalement cette variable. Trott & Bergen (2019) ont par exemple construit des histoires écrites se terminant par un énoncé ambigu, pouvant être interprété comme une requête indirecte ou littéralement (e.g. « Il fait froid ici » peut être interprété comme une demande indirecte d'allumer le chauffage, ou comme un simple commentaire sur la température). Chaque histoire incluait un obstacle à l'accomplissement de cette requête (dans l'exemple précédent, le radiateur était cassé). De plus, chaque histoire était déclinée en deux versions, selon que le locuteur avait connaissance ou non de cet obstacle. Si le locuteur a connaissance de l'obstacle, cela rend improbable qu'il émette une requête : il est par exemple peu probable qu'un locuteur demande à allumer le chauffage s'il sait qu'il ne fonctionne pas. Après la présentation de l'histoire, il était demandé aux participants de sélectionner la paraphrase appropriée de l'énoncé ambigu parmi deux propositions, l'une explicitant la requête en utilisant une forme conventionnelle (e.g. « est-ce que tu peux allumer le chauffage ? ») et l'autre véhiculant le sens littéral de l'énoncé (e.g. « il fait vraiment froid, dommage que le chauffage soit cassé ! »). Trois expériences successives avec ce paradigme confirment ainsi que les participants s'appuient bien sur les états mentaux qu'ils attribuent au locuteur pour interpréter ses énoncés puisque, (1) ils sélectionnent plus fréquemment les paraphrases explicitant la requête lorsque le locuteur n'a pas connaissance de l'obstacle, (2) la propension des participants à moduler leur interprétation selon les états mentaux qu'ils attribuent au locuteur est variable et peut être prédite par leurs compétences en théorie de l'esprit, telles qu'évaluées via une autre tâche expérimentale (la *Short Story Task* de Dodell-Feder et al., 2013), et (3) les mêmes effets sont observés lorsqu'il est explicitement demandé aux participants de juger si le locuteur émet une requête ou non. Dans une série d'expériences et d'analyses supplémentaires, Trott & Bergen (2020) précisent et nuancent leur proposition, en montrant que l'association entre la compréhension de requêtes indirectes et la théorie de l'esprit est modulée par le caractère explicite ou implicite des informations qui sont disponibles. Ils proposent ainsi que « *la théorie de l'esprit pourrait être recrutée de façon flexible et en fonction des besoins [...], les individus*

*exploitant des ressources cognitives et des sources d'informations variées pour reconstruire les intentions de communication d'un locuteur à partir d'un input lacunaire, souvent sous-spécifié, et la variabilité interindividuelle pourrait influencer la fréquence à laquelle des stratégies particulières sont recrutées »* (p. 19, traduction personnelle).

### **3.3. Nuances théoriques et enjeux méthodologiques**

Les liens entre pragmatique et théorie de l'esprit sont soutenus par un ensemble de données, mais il apparaît donc aussi que ces liens sont complexes et modulables. La mobilisation des aspects plus complexes de la théorie de l'esprit, impliquant la représentation explicite d'états mentaux de haut niveau du locuteur ne serait ainsi pas toujours nécessaire ; une prise en compte plus basique de la compatibilité entre une interprétation potentielle et la perspective du locuteur, moins complexe, pourrait être suffisantes dans certains cas (Jary, 2013; Kissine, 2015). Par exemple, comprendre l'ironie nécessiterait en général la représentation d'états mentaux de second ordre, quand des états mentaux de premier ordre seulement seraient nécessaires pour comprendre une métaphore (Happé, 1993). Sans que cela ne semble remettre fondamentalement en question la théorie de la pertinence, Kissine (2015) développe ainsi un modèle dans lequel le recrutement de ces habiletés complexes serait dépendant du contexte et où différentes stratégies d'interprétations d'un énoncé pourraient coexister ; leur sélection serait opérée par une compétence *métacognitive*, en fonction du contexte, des informations disponibles et accessibles, mais également de différences interindividuelles (et notamment la capacité ou non à opérer ces processus complexes, comme nous le verrons dans les chapitres suivants). Soutenant cette hypothèse, Deliens et al. (2017) ont ainsi montré que percevoir l'ironie d'un énoncé pourrait être coûteux par rapport à des énoncés littéraux, suscitant plus d'erreurs et des temps de réponse plus importants. Cela nécessitait en effet de prendre en compte les représentations du locuteur afin de détecter que le contenu littéral de l'énoncé n'était pas cohérent avec celles-ci. Cependant, lorsque la nature ironique de ces énoncés ne reposait pas uniquement sur des informations contextuelles, mais qu'elle était également marquée par des indices prosodiques, les participants ne montraient plus de signes d'efforts particulier par rapport à la condition littérale, indiquant qu'ils pouvaient s'appuyer sur ces indices immédiatement accessibles, sans s'engager dans un raisonnement cognitivement coûteux et explicite sur les représentations de haut niveau du locuteur.

Les interrelations entre la compréhension pragmatique et la théorie de l'esprit (ou plus exactement les différents processus que cela recouvre) sont donc complexes, ce qui questionne la proposition plus élémentaire de la théorie de la pertinence, dans laquelle la compréhension pragmatique est un simple sous-module de la théorie de l'esprit. Nous aurons l'occasion dans le [chapitre 3](#) de voir que cette discussion est notamment alimentée par les performances pragmatiques des personnes autistes. Par ailleurs, cette hypothèse peut poser un problème méthodologique, puisqu'elle conduit à considérer

des mesures pragmatiques comme constituant des mesures de théorie de l'esprit. C'est typiquement le cas des *Strange Stories* (Happé, 1994; O'Hare et al., 2009), qui sont classiquement utilisées comme une mesure de théorie de l'esprit avancée. Comme décrit dans l'*encart 1*, dans chacune des histoires introduites dans ce test, un personnage dans une situation donnée produit un énoncé qui est littéralement faux, pour des raisons variées. Il est donc nécessaire de se représenter les états mentaux du locuteur (connaissances, intentions, croyances) pour déterminer pourquoi cette personne a produit cet énoncé, ce qui explique que ces situations soient utilisées pour évaluer la théorie de l'esprit. Cependant, il apparaît aussi nettement que la réussite à cette épreuve nécessite d'inférer les intentions *de communication* du locuteur, et ce qui motive son énoncé, et qu'il s'agit aussi en cela d'une épreuve de compréhension pragmatique. Cela ne pose pas nécessairement de problème en tant que tel, mais cela rend impossible l'utilisation de cette tâche précisément pour étudier l'articulation entre pragmatique et théorie de l'esprit, puisque ces deux construits sont confondus au sein de cet outil. La proximité de ces construits suscite ainsi des confusions théoriques qui entravent l'étude de leurs interrelations (voir Bosco et al., 2018)

Par ailleurs, une autre barrière méthodologique importante (qui a aussi des implications directes en clinique, comme nous le verrons au *chapitre 3*) affecte cette tâche comme la plupart des tâches de théorie de l'esprit, en cela qu'elles engagent fortement les compétences verbales des participants. C'est également le cas de tâches évoquées ci-dessus telles que les tâches classiques de fausses croyances de premier ordre (Wimmer & Perner, 1983) ou de deuxième ordre (Perner & Wimmer, 1985), des *Strange Stories* (Happé, 1994), du test RME (Baron-Cohen et al., 2001) ou de la Short Story Task (Dodell-Feder et al., 2013) par exemple. Le plus souvent, ces tâches nécessitent l'introduction d'un personnage qui fait l'expérience de situations relativement complexes, décrite verbalement dans des histoires orales ou écrites, puis des questions verbales sont posées et la réponse attendue du participant est souvent verbale également. Le cas du test RME est un peu différent puisque le stimulus est totalement non-verbal (une photographie des yeux d'un acteur) mais le format de réponse consiste à choisir une étiquette verbale écrite parmi 4 pour décrire ses états mentaux, et les étiquettes verbales utilisées sont parfois elles-mêmes complexes et peu fréquentes ; cela explique que la maîtrise de la langue ait une influence certaine sur les performances à cette tâche (Prevost et al., 2014). Bien souvent, ces tâches incluent une condition contrôle ayant pour but de capturer la part des capacités langagières et du raisonnement en général sur les performances dans la tâche, ce qui peut fournir une solution acceptable dans plusieurs contextes. Cependant, dans le cas d'une utilisation de ces tâches pour étudier les liens entre inférences pragmatiques et théorie de l'esprit, cela peut constituer une variable confondante importante puisque les épreuves pragmatiques mettent nécessairement en jeu du matériel verbal également ; une corrélation entre les performances à une tâche de compréhension pragmatique et une tâche de théorie de l'esprit pourrait être due à l'implication de la théorie de l'esprit mais également des compétences verbales dans les deux tâches. Au-delà de ces aspects méthodologiques, le langage et la théorie de l'esprit entretiennent également des liens

développementaux forts (voir par exemple de Villiers, 2007; Durrleman et al., 2019), qui rendent d'autant plus complexe l'étude de l'articulation entre langage, pragmatique et théorie de l'esprit. Privilégier des tâches de théorie de l'esprit non-verbales dans ce contexte paraît donc important, mais apparaît comme un choix encore peu fréquent.

## Vers une approche développementale

Ce chapitre a permis d'établir le cadre général de cette thèse, dans lequel les êtres humains ne communiquent pas en transmettant de l'information à l'aide d'un code, mais en s'influençant les uns les autres, notamment avec des énoncés verbaux qu'ils utilisent comme des indices de leurs intentions. Le décodage linguistique d'un énoncé n'est donc pas suffisant pour communiquer, comme observé dans deux phénomènes prototypiques que sont les implicatures scalaires et les métaphores. Accéder aux intentions de communication nécessite ainsi des inférences pragmatiques, qui sont opérées à un certain coût cognitif, modulé par différents facteurs contextuels ou intrinsèques. Ces processus s'appuient sur l'habileté plus générale de l'être humain à raisonner sur les états mentaux d'autrui, dans une interface complexe. Ces modèles théoriques sont adossés à de nombreux travaux expérimentaux portant sur l'individu mature, expert sur le plan linguistique et communicationnel, mais ont aussi largement bénéficié de travaux développementaux (Karmiloff-Smith, 1998). L'observation de retards développementaux dans l'opération d'une inférence pragmatique, par rapport au traitement littéral d'un énoncé, est par exemple un marqueur fort du coût cognitif de cette opération (Noveck, 2018). Nous pouvons désormais nous tourner vers l'approche développementale qui est au cœur de cette thèse, en examinant comment se développent ces inférences pragmatiques chez l'enfant.

## Chapitre 2 • Le développement des inférences pragmatiques

### 1. Généralités

Avant d'examiner l'état des connaissances spécifiques au développement des métaphores (*partie 2*) et des implicatures scalaires (*partie 3*), il sera utile de situer ces travaux au sein d'une littérature plus large sur les inférences pragmatiques ; celle-ci révèle l'apparent paradoxe des inférences pragmatiques, qui constituent à la fois les aspects les plus précoces et les plus tardifs du développement langagier humain, et montre que ce développement s'appuie sur des fondations cognitives multiples.

#### 1.1. Fondations précoces et développements tardifs

Telles que définies au *premier chapitre*, les compétences pragmatiques apparaissent précocement chez l'enfant, non seulement *avant* le développement des compétences langagières formelles, mais aussi comme un *moteur* de ces acquisitions. En effet, l'enfant montre dès la naissance des habiletés à communiquer, d'abord dans une relation dyadique avec l'adulte, puis, à partir d'environ 9 mois, dans une communication dite triadique qui intègre le monde extérieur (pour une revue voir Stephens & Matthews, 2014), avec selon certains auteurs une prédisposition innée à détecter la communication ostensive dirigée vers l'enfant (Csibra, 2010). S'appuyant sur une compréhension rudimentaire d'autrui comme poursuivant des buts, observable dès 6 mois (Woodward, 1999), les jeunes enfants distinguent progressivement différents types d'intentions chez l'autre. À 9 mois, ils réagissent par exemple différemment à un adulte qui ne réalise pas une action parce qu'il ne le *veut* pas ou parce qu'il ne le *peut* pas (Behne, Carpenter, Call, et al., 2005), manifestant ainsi une compréhension d'intentions même non-réalisées. À 18 mois, ils peuvent en outre imiter l'action qu'un expérimentateur voulait mais avait échoué à réaliser (Meltzoff, 1995). Plus spécifiquement, les enfants présentent même une compréhension précoce des *intentions de communication* des autres. Dès 14 mois, ils distinguent par exemple des regards ou des pointages réalisés avec ou sans intention manifeste de communiquer, les utilisant pour localiser un objet seulement lorsqu'ils sont associés à une telle intention (Behne, Carpenter, & Tomasello, 2005). Sur le versant expressif, dès 18 mois, lorsqu'ils font par exemple des demandes, les enfants ne communiquent pas seulement pour obtenir des objets, mais aussi pour que leur intention même de communiquer soit reconnue, montrant ainsi une communication ostensive-inférentielle au sens de Grice, et pas seulement utilitaire (Grosse et al., 2010).

Observable avec des comportements non-verbaux, cette habileté précoce à inférer les intentions de communication d'autrui intervient également dans les premières

utilisations du langage et guide son acquisition. En effet, par principe, la confrontation à un mot nouveau nécessite de s'appuyer sur d'autres informations pour déterminer son sens ou son référent, en inférant les intentions de communication du locuteur. Bannard et Tomasello (2012) ont par exemple montré que, bien qu'ils étaient capable d'associer un mot avec un référent depuis beaucoup plus jeune, des enfants de 24 mois n'utilisaient ce mot pour inférer les intentions de communication d'autrui que si ce mot avait été acquis dans un contexte d'interaction. Il est par ailleurs bien établi qu'en entendant un mot nouveau, les jeunes enfants excluent des objets familiers ou qui ont déjà été nommés comme référents potentiels (e.g. E. V. Clark, 1990; Markman & Wachtel, 1988; Tomasello & Akhtar, 1995). Cette observation pourrait témoigner d'un raisonnement contrefactuel précoce, pragmatique par nature : si le locuteur avait voulu désigner l'objet connu, il aurait utilisé le mot connu. Un mot inconnu désigne alors probablement un objet inconnu. Cette hypothèse est soutenue par un large corpus expérimental (pour une revue voir Grassmann, 2014) qui montre que le jeune enfant infère le sens d'un mot en exploitant un ensemble d'informations pragmatiques telles que le regard (e.g. Baldwin, 1991), les gestes et notamment les pointages (e.g. Grassmann & Tomasello, 2010), le contexte verbal du mot en question (e.g. Saylor et al., 2002), la nature fiable ou non du locuteur (e.g. Birch et al., 2008; Sabbagh & Baldwin, 2001) et même ses croyances (Houston-Price et al., 2011).

Mais malgré ces dispositions précoces, on considère classiquement que les enfants rencontrent jusque tardivement des difficultés à inférer précisément les intentions de communication des locuteurs lorsqu'elles divergent du sens linguistique de l'énoncé (Airenti, 2017; Falkum & Köder, 2020; Nippold, 1998). C'est typiquement dans le cas avec les métaphores ou les implicatures scalaires, comme nous le verrons dans la suite de ce chapitre. Ackerman (1981) montre ainsi que des enfants peuvent détecter des violations des maximes gricéennes à l'âge de 6/7 ans, mais qu'ils ne les utilisent correctement pour inférer les intentions de communication du locuteur que vers 8 ou 9 ans. Parmi les phénomènes les plus complexes à appréhender, l'ironie commencerait à être comprise vers 5 ou 6 ans mais continuerait de se développer tardivement, jusque dans l'adolescence (Filippova, 2014; Filippova & Astington, 2008), de même que l'humour (Hoicka, 2014), qui s'appuie en outre sur des représentations conventionnelles et culturelles fines.

## **1.2. Bases cognitives du développement pragmatique**

Le développement pragmatique est donc modélisé comme s'appuyant sur le développement de la théorie de l'esprit de l'enfant, pour des raisons théoriques (*chapitre 1*), qui sont aussi étayées par des données expérimentales. Premièrement, un parallèle peut être établi entre les trajectoires développementales de ces deux compétences, dont on observe à la fois des émergences précoces, voire innées, et des développements tardifs jusqu'à l'âge adulte (voir ci-dessus et *l'encart 2* sur le développement plus particulier de la compréhension des fausses croyances).

### **Encart 2 : Développement de la compréhension des fausses croyances**

À l'issue des travaux de Wimmer et Perner (1983), démontrant une compréhension des fausses croyances de premier ordre qui émergerait vers 4-5 ans (voir [encart 1](#)), ce jalon développemental a été largement confirmé (Wellman et al., 2001). La littérature sur la théorie de l'esprit s'est également développée autour de cette situation paradigmatique, dans deux directions opposées. Il a ainsi été montré que la théorie de l'esprit continuait à se développer et à se complexifier après l'âge de 4 ans, à l'âge scolaire et au-delà. Si les résultats précédents portaient sur des croyances de premier ordre (Jean pense que X), des croyances de deuxième ordre peuvent également être conceptualisées (Marie croit que Jean pense que X), et sont plus complexes à traiter. Les fausses croyances de deuxième ordre sont typiquement comprises plus tard par les enfants, vers l'âge de 7 ans (Perner & Wimmer, 1985). Des développements plus complexes (par exemple des fausses-croyances de troisième ordre) peuvent être observés plus tard encore, jusqu'à la fin de l'adolescence et au début de l'âge adulte (Blakemore, 2008; Bosco et al., 2014; Valle et al., 2015).

À l'inverse, des signes plus précoces de théorie de l'esprit peuvent être observés avant l'âge de 4 ans au moyen de paradigmes plus subtils (Frith & Frith, 2003). Par exemple, des scénarii classiques comme celui de Maxi (cf. [chapitre 1](#)) ont été adaptés dans une tâche d'eye-tracking. Avec ce paradigme, Southgate et al. (2007) ont ainsi démontré que bien avant de pouvoir répondre correctement à des questions, les enfants dès l'âge de 2 ans regardent de préférence l'endroit où le personnage *pense* que le chocolat se trouve, plutôt qu'à l'endroit où eux-mêmes savent qu'ils se trouvent. Dans cette tâche dite de théorie de l'esprit *implicite*, ils manifestent ainsi une compétence précoce pour se représenter les états mentaux d'un personnage, et à les utiliser pour prédire leur comportement, avant d'être capables de répondre à des questions explicites. Similairement, Onishi et Baillargeon (2005) ont utilisé un paradigme de transgression des attentes, montrant que des enfants dès l'âge de 15 mois étaient surpris lorsqu'un adulte se comportait autrement que selon ses croyances, vraies ou fausses, témoignant d'une représentation des (fausses) croyances d'autrui plus précoces encore.

Par ailleurs, certaines études ont mis en relation les différences interindividuelles dans les inférences pragmatiques et la théorie de l'esprit au cours du développement. Cette association a particulièrement été étudiée autour de l'ironie, dont la compréhension nécessiterait de comprendre qu'un locuteur *pense* que son destinataire *sait* qu'il n'est pas littéral, et de saisir *l'attitude* du locuteur vis-à-vis de son énoncé, c'est-à-dire qu'elle nécessiterait un raisonnement récursif complexe sur les états mentaux du locuteur. Des corrélations fortes entre des tâches de compréhension de l'ironie et de théorie de l'esprit de deuxième ordre ont ainsi pu être observées dans différentes études (e.g. Nilsen et al., 2011). Filippova & Astington (2008) ont par exemple évalué des enfants de 5, 7 et 9 ans, à l'aide d'histoires verbales qui leur étaient lues, associées à des supports visuels, et dans lesquelles une remarque ironique était faite à un personnage. Après chaque histoire, une série de questions étaient posées aux participants, renseignant sur la compréhension de l'histoire (comme mesure contrôle), la détection de l'ironie et la compréhension des intentions du locuteur, aboutissant à un score composite de compréhension de l'ironie. Ils ont en outre proposé des épreuves

de théorie de l'esprit complexe, évaluant notamment la représentation d'états mentaux de deuxième ordre, ainsi que des épreuves de langage (compréhension lexicale), de mémoire de travail et de perception de la prosodie émotionnelle. Outre un effet développemental clair de la compréhension de l'ironie, les résultats montraient une forte corrélation entre cette compétence et la théorie de l'esprit. Toutefois, l'ensemble des variables mesurées étaient corrélées entre-elles, et corrélées avec l'âge. La compréhension lexicale en particulier, était également fortement associée à l'âge et à la compréhension de l'ironie. Afin de déterminer la contribution relative de ces différents facteurs, les auteurs ont ajusté des modèles de régressions linéaires hiérarchiques, qui ont révélé qu'au-delà de l'âge, de la mémoire de travail et de la perception de la prosodie, la compréhension du langage et la théorie de l'esprit apportaient des contributions significatives et distinctes à la variance en compréhension de l'ironie. Ainsi, cette étude confirmait l'influence spécifique de la théorie de l'esprit sur la compréhension pragmatique, mais mettait aussi en évidence l'importance d'autres facteurs, et notamment des compétences langagières formelles.

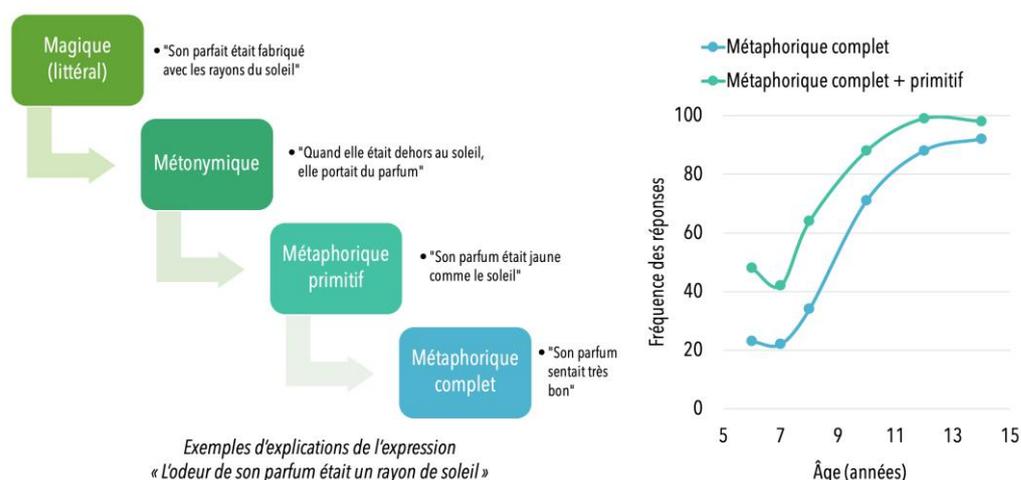
Ces résultats font échos à des propositions théoriques issues de l'étude de fonctionnements atypiques, et notamment de l'autisme, qui postulent que la théorie de l'esprit et les compétences langagières formelles seraient des socles cognitifs importants de la compréhension pragmatique (cf. Happé, 1993; Norbury, 2004 ; propositions détaillées au [chapitre 3](#)). L'étude des différences interindividuelles au sein du développement typique est toutefois plus récente, comme le montrent Matthews et al. (2018) dans une large revue de la littérature. Cette revue confirme que plusieurs travaux mettent en évidence des associations entre les habiletés pragmatiques des enfants en général, et leur théorie de l'esprit, leurs compétences langagières formelles, mais aussi leurs fonctions exécutives. Cependant, les auteurs pointent également d'importantes limites de cette littérature. Ils relèvent notamment que ces associations ont souvent été observées dans des études distinctes, et que l'étude de Filippova & Astington (2008) rapportée ci-dessus demeure en cela une exception : peu d'études ont mesuré *conjointement* les influences respectives de différents prédicteurs potentiels. De plus, la structure factorielle des fonctions exécutives demeure peu claire, ce qui complexifie encore l'observation de ce type d'association, et la comparabilité des résultats. Pour ce qui est de la théorie de l'esprit, relativement mieux circonscrite, cette revue pointe la faible qualité de la plupart des tâches utilisées, qui génère une variance limitée, ce qui limite les associations pouvant être mises en évidence. À cela peut être ajoutée une limite de nombreuses tâches de théorie de l'esprit, qui mobilisent fortement les compétences langagières (comme développée au [premier chapitre](#)). Cette variable confondante constitue une limitation particulièrement critique pour l'étude des rôles respectifs du langage et de la théorie de l'esprit dans le développement des inférences pragmatiques.

Dans la suite de ce chapitre, nous examinerons plus spécifiquement les trajectoires développementales des métaphores ([partie 2](#)) et des implicatures scalaires ([partie 3](#)), ainsi que les possibles bases cognitives qui les sous-tendent.

## 2. Développement des métaphores

### 2.1. Trajectoire développementale de la compréhension des métaphores

Le développement des métaphores a fait l'objet d'une importante littérature durant les 50 dernières années. Dans une étude de référence, Winner et al. (1976) ont présenté à des enfants de 6 à 14 ans des expressions métaphoriques telles que « *L'odeur de son parfum était un rayon de soleil* » et ont demandé aux participants de les expliquer verbalement, ou de sélectionner une interprétation correcte parmi un ensemble de propositions. Les propositions d'interprétation étaient formulées selon le modèle développemental des auteurs, incluant, du plus simple au plus complexe, i) une interprétation dite *magique*, c'est à dire une interprétation littérale de l'expression, rendue plausible dans un monde magique (voir des exemples d'interprétations en figure 2-1, à gauche), ii) une interprétation *métonymique*, c'est-à-dire une reformulation mettant en relation la signification littérale des deux termes de l'expression de façon plausible dans le monde réel, iii) une interprétation *métaphorique primitive*, c'est-à-dire utilisant une caractéristique accessoire du véhicule pour l'appliquer à la cible, mais sans établir de comparaison entre un domaine physique et un domaine psychologique, ou entre différentes modalités sensorielles ; et enfin iv) une interprétation *métaphorique complète*, c'est-à-dire que la caractéristique centrale du véhicule est correctement transférée à la cible de la métaphore, réalisant une comparaison implicite entre un domaine physique et un domaine psychologique, ou entre différentes modalités sensorielles. Par ailleurs, les réponses libres des participants dans la condition 'explication verbale' étaient catégorisées selon ces 4 types de réponses.



**Figure 2-1 :** Stades d'interprétation des métaphores selon Winner et al. (1976) avec des exemples d'interprétation de la métaphore "L'odeur de son parfum était un rayon de soleil" (à droite), et la fréquence des réponses de type métaphorique complet ou primitif, en modalité d'explication verbale libre, en fonction de l'âge (valeurs extraites de la publication originale)

Winner et al. (1976) ont ainsi observé une nette progression des performances avec l'âge, dans les deux modalités, qui correspondait aux stades développementaux décrits

ci-dessus (voir figure 2-1, à droite pour l'évolution de la fréquence des interprétations métaphoriques libres). Par ailleurs, la modalité de réponse libre révélait aussi également des interprétations métaphoriques inappropriées, dans lesquelles les enfants connectaient bien deux concepts de domaines différents, mais en sélectionnant des propriétés inappropriées (par exemple, « Son parfum avait une odeur marrante »), et qui précédaient les réponses correctes. Ce n'est que vers 10 ans que les enfants privilégiaient clairement les propositions de réponse métaphoriques complètes. En modalité d'explication libre, toutefois, ils demeuraient à cet âge susceptibles de compléter leur explication par une interprétation métaphorique primitive ou métonymique. Dans cette modalité, les réponses métaphoriques complètes ne devenaient majoritaires qu'à l'orée de l'adolescence (figure 1-2, droite).

Par ailleurs, en plus d'indiquer un développement progressif et tardif clair, cette étude mettait en évidence deux facteurs modulant cet effet. Premièrement, les performances des enfants apparaissaient meilleures lorsqu'ils devaient sélectionner une réponse parmi plusieurs propositions, que lorsqu'ils devaient expliquer librement leur interprétation. Par ailleurs, différents types de métaphores étaient proposées : des métaphores mettant en relation deux domaines sensoriels distincts, tels que l'exemple décrit ci-dessus (mettant en lien la modalité olfactive et la modalité visuelle) ou des métaphores mettant en relation un concept psychologique avec un concept physique (par exemple décrire un gardien de prison comme étant 'dur comme de la pierre'). Les métaphores faisant intervenir des concepts psychologiques étaient associées à des performances plus faibles que les métaphores dites sensorielles. Nous verrons dans la suite de ce chapitre que ces deux facteurs (format de réponse et type de métaphore) ont fait l'objets d'expérimentations ultérieures.

L'étude de Winner et al. (1976) est représentative d'une série de travaux qui confirment le développement tardif de la compréhension des métaphores (voir aussi Asch & Nerlove, 1960, et les revues de Nippold, 1998; Winner, 1997), mais qui se confrontent également à plusieurs articles rapportant des manifestations beaucoup plus précoces d'une certaine maîtrise de ces phénomènes. En effet, des exemples de productions pouvant être considérées comme métaphoriques ont ainsi été attestés chez des enfants dès 2 ans (e.g. Carlson & Anisfeld, 1969; Gentner, 1988; Harris, 1982). Ces exemples sont toutefois difficilement distinguables d'expressions de faire-semblant, ou de surgénéralisations, fréquemment observées chez le jeune enfant (e.g. un enfant qui regarde du sable s'écouler d'un tamis et commente « Il pleut »<sup>10</sup>). La recherche d'une compréhension précoce des métaphores a plus récemment fait l'objet d'un regain d'attention, appuyé sur les travaux évoqués à la [section 1.1](#), et montrant que de nombreuses compétences pragmatiques sont précocement fonctionnelles chez l'enfant. Pouscoulous (2011, 2014) propose ainsi que les enfants puissent être compétents pour comprendre les métaphores de façon précoce, mais que ces compétences soient masquées par les paradigmes expérimentaux classiquement

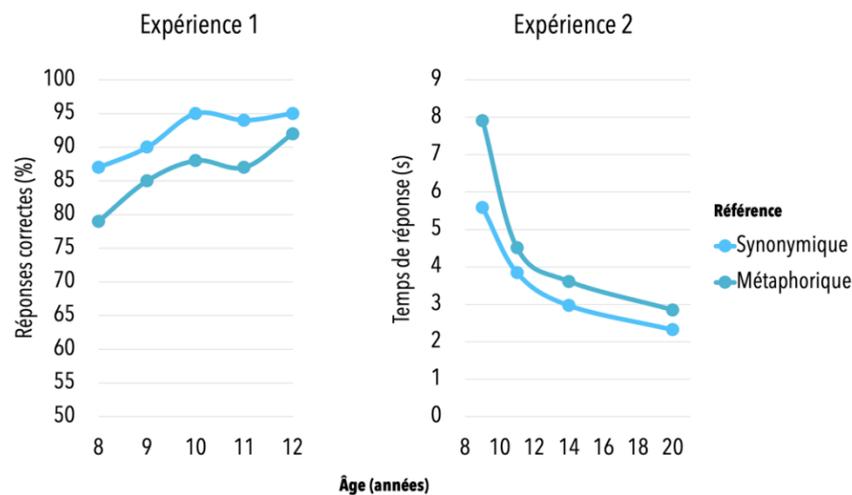
---

<sup>10</sup> Traduction d'un exemple rapporté par Pouscoulous et al. (2014)

utilisés. Quatre facteurs non-pragmatiques, également étudiés chez l'adulte (voir [chapitre 1](#)) pourraient ainsi expliquer les mauvaises performances habituelles des enfants dans la littérature : (1) l'utilisation de métaphores avec peu de contexte, contrairement aux utilisations naturelles de ces expressions ; (2) les connaissances lexicales et encyclopédiques souvent complexes qui doivent être exploitées pour interpréter une métaphore, expliquant par exemple que les métaphores sensorielles soient maîtrisées plus précocement que les métaphores dites psychologiques ; (3) l'utilisation de métaphores souvent conventionnelles voire d'idiomes, nécessitant un apprentissage en tant que tel, par rapport à des métaphores nouvelles qui peuvent être traitées sans un tel apprentissage ; et (4) la complexité de la tâche expérimentale employée, en termes de charge linguistique générale des stimuli à traiter, et de compétences métalinguistiques mobilisées dans des tâches de jugements, ou d'explications verbales. Pouscoulous et Tomasello (2020) ont ainsi conçu une tâche neutralisant ces facteurs, dans laquelle un expérimentateur demandait à des enfants de leur donner un objet en s'y référant à l'aide d'une métaphore (par exemple « *Donne-moi le chien avec les chaussures marrons* », désignant un chien avec des pattes marrons). Ce mode de réponse non-verbal limitait ainsi la charge métalinguistique de la tâche. De plus, chaque métaphore était non-conventionnelle, mobilisait des mots et objets fréquents, était basée sur des similarités physiques directement observables, et ne créait pas de situation ambiguë, c'est-à-dire que l'expression ne pouvait pas littéralement désigner un autre référent (dans l'exemple précédent, l'objet en compétition avec la cible n'était pas un chien avec des chaussures littéralement marrons, mais un chien avec un dos marron). Par ailleurs, la connaissance du vocabulaire impliqué dans les métaphores et dans les objets utilisés était évaluée, en compréhension et en production, à l'issue de la tâche expérimentale. L'évaluation avec cette tâche de 24 enfants de 3 ans a révélé que les enfants sélectionnaient l'objet cible plus fréquemment que le hasard (73% de réponses correctes au total), malgré une certaine variabilité inter-items. Aucune corrélation n'était toutefois observée entre la compréhension des métaphores et la connaissance du vocabulaire impliqué. Ce résultat démontre ainsi l'existence d'une compréhension précoce des métaphores, bien que partielle, dès l'âge de 36 mois.

D'autres travaux ont cependant continué de confirmer et de préciser l'amélioration de la compréhension des métaphores au-delà de ces émergences précoces, à l'âge scolaire. Noveck et al. (2001) ont par exemple proposé à des enfants de 8 à 12 ans un paradigme de références métaphoriques, sur la base de celui de Gibbs (1990), dans lequel des vignettes écrites sont présentées, introduisant un référent dans la première partie de l'histoire (par exemple, des élèves qui vont à la piscine), puis rappelé à la fin de l'histoire par une référence. Chaque vignette est déclinée dans une paire d'histoires, à la fin desquelles la référence est ou bien métaphorique (par exemple 'les crapauds'), ou synonymique (ici, 'les enfants'). Après chaque histoire, une question fermée permet d'établir si la référence a bien été établie. Noveck et al. (2001) montrent ainsi que, si la compréhension des références augmente progressivement avec l'âge, les références métaphoriques suscitent plus d'erreurs que les références contrôles synonymiques, et

ce jusqu'à 12 ans, où l'écart semble seulement se refermer (voir figure 2-2, gauche). Dans une deuxième expérience, les auteurs adaptent ces stimuli dans une tâche de lecture informatisée, qui permet de mesurer le temps nécessaire pour lire les expressions référentielles d'intérêt. Les données révèlent alors que les références métaphoriques (interprétées correctement) sont lues plus lentement que les références synonymiques, y compris chez des adultes qui ont également été testés dans cette expérience (figure 2-2, droite). Cette différence de temps de lecture était particulièrement importante à 9 ans (plus de 2000 ms), et relativement stable au-delà (environ 650 ms à 11 et 14 ans, et 530 ms chez les adultes), confirmant le coût temporel associé aux références métaphoriques, y compris chez des enfants. Par ailleurs, dans cette seconde expérience, les proportions de réponses correctes montraient qu'avec l'âge, non seulement les références métaphoriques n'étaient pas plus difficiles à établir que les réponses synonymiques, mais les adultes montraient même de *meilleures* performances dans la condition métaphorique que dans la condition contrôle, en lien avec les bénéfices associés aux métaphores évoqués dans le [premier chapitre](#).



**Figure 2-2 :** Taux de réponses correctes (gauche, expérience 1) et temps de réponse (droite, expérience 2) en fonction de l'âge à la tâche de références métaphoriques proposée par Noveck et al. (2001, données extraites de la publication originale ; la moyenne d'âge du groupe d'adultes a été utilisée pour l'expérience 2)

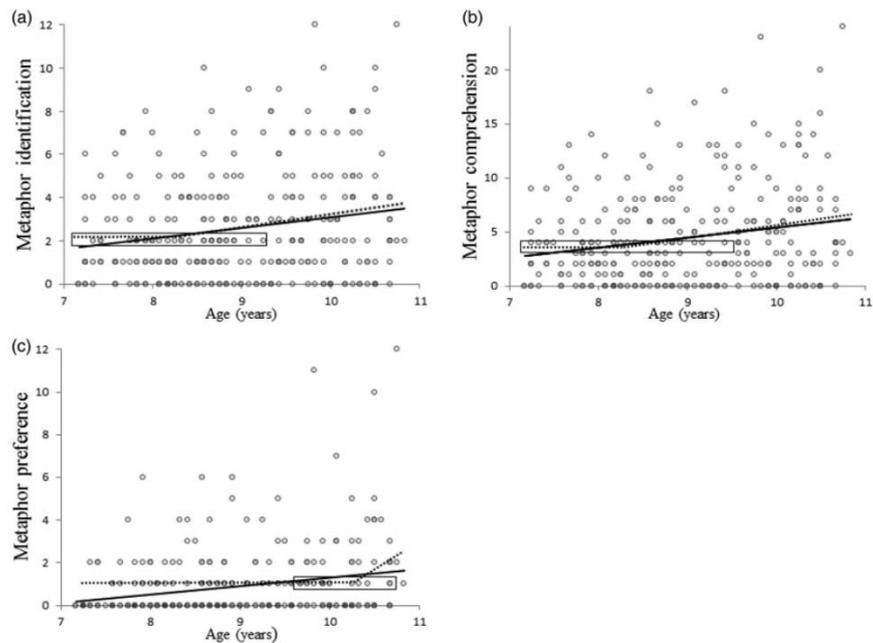
Seigneuric et al. (2016) ont adapté ce paradigme en posant une question ouverte plutôt que fermée. Ils ont par ailleurs fait varier la congruence sémantique entre le véhicule et la cible, en proposant des métaphores dont les éléments appartenait aux mêmes catégories générales (entre deux entités humaines, ou entre deux objets, donc relativement proches) ou non (entre une entité humaine et une entité objet, donc relativement distantes). Des enfants de 8 à 10 ans ont été évalués avec ce paradigme, répartis sur la base de leurs performances en compréhension de texte en un groupe de bon compreneurs, et un groupe de mauvais compreneurs. Les deux groupes étaient par ailleurs appariés en termes de décodage du langage écrit, de vocabulaire et d'intelligence non-verbale. Bien que les effets liés à l'âge ne soient pas rapportés, les résultats de cette étude confirment et étendent ceux de Noveck et al. (2001), en montrant que l'écart des performances entre la condition contrôle et la condition

métaphore est particulièrement important pour les faibles compreneurs, et plus restreint chez les bons compreneurs, et que dans le premier groupe, l'incongruence sémantique majore les difficultés à comprendre les métaphores. De plus, le format de réponse ouvert a permis une analyse qualitative des erreurs, montrant qu'une part non négligeable d'entre-elles constituaient des interprétations littérales du référent, en particulier chez les mauvais compreneurs (41% des erreurs, contre 21% chez les bons compreneurs ; résultats recalculés à partir des données de la publication originale).

Enfin, Van Herwegen et al. (2013) ont également utilisé une tâche similaire mais proposée à l'oral, avec un support visuel imagé. Afin d'évaluer si la référence métaphorique avait bien été comprise, les participants devaient répondre à une question en sélectionnant une image parmi 3, représentant une interprétation métaphorique (correcte), une interprétation littérale du référent, et un distracteur. Cette étude n'incluait pas de condition contrôle, mais a montré une progression de la compréhension des métaphores avec l'âge, chez un groupe d'enfants et adolescents de 3 à 17 ans. Les auteurs précisent toutefois que leur échantillon initial n'incluait que des enfants de 6 à 17 ans, et qu'il a été nécessaire d'inclure des enfants plus jeunes, de 3 à 6 ans, pour obtenir une variabilité liée à l'âge sur certains items qui étaient déjà largement réussis à 6 ans.

Plus récemment, l'équipe de Willinger (Deckert et al., 2019; Willinger et al., 2019) a adapté la tâche dite des triades métaphoriques (Metaphor Triad Task, MTT, Kogan & Chadrow, 1986) dans laquelle 3 mots ou expressions sont présentées aux participants (par exemple : grand-père, rocking chair, vieil arbre), avec lesquels 3 paires de mots peuvent être formés. Crucialement, une de ces associations est métaphorique (ici, grand-père et vieil arbre), tandis que les deux autres sont non-métaphoriques. Pour chaque triade, il est demandé aux participants de former la meilleure association possible, puis toutes les autres associations possibles, et d'expliquer chaque association proposée. Cette consigne permet d'établir 3 scores pour chaque triade : (1) un score d'identification, selon si la relation métaphorique a été identifiée ou non ; (2) un score d'explication, selon si cette relation était correctement justifiée ou non ; et (3) un score de préférence, selon si la relation métaphorique était citée comme la meilleure association possible ou non. Willinger et al (2019) ont utilisé cette tâche avec des enfants de 7, 9 et 11 ans, et ont montré une nette amélioration de la compréhension entre 9 et 11 ans, et de moindres différences d'âge entre 7 et 9 ans. Prolongeant ce travail, Deckert et al. (2019) ont testé 298 enfants de 7 à 11 ans afin d'étudier les liens entre les performances et l'âge, cette fois utilisé comme prédicteur continu. Leurs analyses ont mis en évidence que des modèles non-linéaires étaient plus appropriés aux données que des modèles linéaires, avec des âges pivot avant lesquels les performances ne progressaient pas et après lesquels les scores augmentaient (voir figure 2-3). Les âges pivots étaient différents selon scores considérés, avec un pivot à 8,2 ans pour l'identification et l'explication des métaphores, et un deuxième pivot à 10,2 ans, après lequel la préférence pour l'association métaphorique augmentait. Pour ces auteurs, cela constitue un argument en faveur d'un changement qualitatif et pas

seulement quantitatif dans la *nature* des processus qui permettent aux enfants de comprendre les métaphores, au cours de l'âge scolaire. Ils lient cette modification aux importants changements qui caractérisent cette période du développement, à la fois au niveau neurologique (Blakemore, 2008), et au niveau comportemental avec une complexification des relations sociales. Nous verrons dans la section suivante que ce changement qualitatif pourrait également être marqué par un changement dans les prédicteurs cognitifs de la compréhension des métaphores.



**Figure 2-3 :** Trajectoires développementales des performances des enfants dans l'étude de Derkert et al. (2019) pour les scores d'identification (a), d'explication (b) et de préférence (c), mettant en évidence les patterns non-linéaires (en pointillés) expliquant mieux les données que des patterns linéaires (traits pleins), avec les intervalles de confiance à 95% pour les âges pivots encadrés (figure extraite de la publication originale)

Enfin, Lecce et al. (2019) ont également étudié le développement des métaphores à l'âge scolaire en introduisant une tâche de métaphores mentales et physiques (*Physical and Mental Metaphors*, PMM), reprenant la distinction introduite dans l'étude de Winner et al. (1976). Cette tâche propose des métaphores sous la forme 'X est Y' dans lesquelles des personnes (X) sont associées à des entités non-humaines (Y), cette association pouvant reposer sur des caractéristiques *physiques* telles que l'apparence ou le comportement (e.g. 'les danseurs sont des papillons'), ou sur des traits *mentaux* ou psychologiques (e.g. 'les soldats sont des lions'). Dans la PMM, ces métaphores sont présentées oralement aux enfants, et il leur est demandé d'en expliquer le sens. Deux mesures différentes sont codées à partir de leurs réponses : si l'explication est correcte ou non (score de réussite) et si l'interprétation proposée s'appuie sur des caractéristiques psychologiques ou non (score d'interprétation). Le matériel est par ailleurs précautionneusement contrôlé sur le plan psycholinguistique, avec des mots fréquents, constituant des métaphores non-lexicalisées et conçues pour être facilement compréhensibles par des enfants. La nature mentale ou physique des métaphores a par

ailleurs été validée dans une étude pilote auprès d'adultes. Lecce et al. (2019) ont évalué à l'aide de cette tâche 217 enfants de 9 à 12 ans. Les résultats confirment que ces métaphores sont bien expliquées, par des justifications plausibles (score de réussite), avec d'ailleurs de meilleures performances sur les métaphores mentales que physiques. Aussi, les auteurs se focalisent sur les scores *d'interprétation*, qui montrent un effet développemental pour les métaphores mentales, et non pour les métaphores physiques, les enfants étant plus enclins à s'appuyer sur des caractéristiques psychologiques pour les expliquer avec l'âge. Cet effet est particulièrement marqué entre 9 et 11 ans. Ces résultats précisent donc les développements à l'œuvre à l'âge scolaire, où des enfants se montrent déjà capable de fournir des explications plausibles pour des métaphores simples, mais où la nature de ces explications évolue, s'appuyant de plus en plus sur des caractéristiques psychologiques. Ce résultat peut donc être mis en parallèle avec les développements observés à cet âge dans la compréhension de ces caractéristiques psychologiques, c'est-à-dire en théorie de l'esprit (cf. [encart 2](#)), ce qui constituait l'objectif principal de cette étude. C'est ce vers quoi nous pouvons désormais nous tourner dans la section suivante.

## **2.2. Prédicteurs cognitifs de la compréhension des métaphores**

Au-delà de l'effet développemental, Lecce et al. (2019) étaient en effet intéressés par les prédicteurs de la compréhension et de l'explication de ces métaphores. Ils ont donc également évalué la théorie de l'esprit de leurs participants avec la tâche des Strange Stories (Happé, 1994), et ont recherché des associations entre les différences interindividuelles en compréhension des métaphores et en théorie de l'esprit. Leurs résultats ont révélé une corrélation entre les scores d'interprétation des métaphores mentales (et non les métaphores physiques) et les performances en théorie de l'esprit, à 9 ans mais pas à 10 ou 11 ans ; cette corrélation demeurait même en contrôlant pour le SES, la compréhension lexicale et la mémoire de travail (ces corrélations avec les scores de réussite n'étaient pas rapportées). Autrement dit, les enfants de 9 ans avec une meilleure théorie de l'esprit avait plus tendance que les autres à s'appuyer sur des traits psychologiques pour expliquer des métaphores mentales, résultat important pour trois raisons. Premièrement, cela confirme la différence de nature entre les métaphores physiques et les métaphores mentales que la PMM exploite. Deuxièmement, cela confirme le rôle important de la théorie de l'esprit pour l'interprétation des métaphores, en lien avec les données obtenues chez l'adulte (voir [chapitre 1](#)) ou pour d'autres phénomènes pragmatiques (voir [section 1.2 du présent chapitre](#)). Avec le même type de tâche, la même équipe a ultérieurement confirmé dans une étude longitudinale de courte durée le lien entre compréhension des métaphores et théorie de l'esprit à 6 mois d'intervalle (Del Sette et al., 2020). La directionnalité de cette association reste en revanche peu claire, puisque cette étude montrait des liens généraux bidirectionnels entre compréhension pragmatique et théorie de l'esprit. Toutefois, les auteurs ont constitué un score composite capturant la propension à s'appuyer sur des traits psychologiques pour interpréter des métaphores mentales spécifiquement par rapport aux métaphores physiques. Ce score était prédictif de la

théorie de l'esprit 6 mois plus tard, et non pas réciproquement, suggérant une influence des compétences pragmatiques sur le développement de la théorie de l'esprit, et non l'inverse. Troisièmement, l'étude de Lecce et al. (2019) suggère que le lien entre compréhension de métaphores et théorie de l'esprit s'amenuise avec l'âge, introduisant la possibilité d'une évolution développementale dans les bases cognitives des métaphores. Ce résultat a par ailleurs récemment été répliqué (Tonini et al., 2023) et étendu à une autre modalité d'évaluation de la compréhension des métaphores, avec le paradigme référentiel utilisé par Noveck et al. (2001). Avec cette tâche, la théorie de l'esprit était associée à la compréhension des références métaphoriques mais pas synonymiques à 8 et 9 ans, mais plus à 10 ans. Les auteurs mettent en relation ce résultat avec le modèle de la modularisation progressive de Karmiloff-Smith (1995), selon laquelle les différents modules cognitifs (ici la compréhension pragmatique et la théorie de l'esprit) se spécialisent et s'individualisent progressivement au cours du développement.

En miroir de cette évolution développementale, l'étude de Decker et al. (2019) avait également investigué les prédicteurs de la compréhension des métaphores, mais en se focalisant sur le rôle des habiletés langagières, plutôt que sur la théorie de l'esprit qu'ils n'avaient pas mesurée. Dans cette étude, les performances à la MTT étaient significativement associées à l'intelligence verbale des enfants, à 9 et à 10 ans, mais pas chez les enfants plus jeunes de 7 et 8 ans. À l'inverse des résultats de Lecce et al. (2019), Deckert et al. (2019) suggèrent plutôt une influence croissante des habiletés langagières sur le traitement des métaphores, indiquant que celles-ci pourraient « *devenir graduellement plus linguistiquement articulées* » (p. 358). Considérés conjointement, ces résultats suggèrent plus globalement que les bases cognitives du traitement des métaphores évolueraient avec l'âge, renforçant l'hypothèse d'un changement *qualitatif* dans le traitement des métaphores à l'âge scolaire.

Ces travaux présentent cependant deux limites importantes qui ont précédemment été évoquées. Premièrement, pour évaluer la théorie de l'esprit, les travaux de Lecce et al. (2019) et de Tonini et al. (2023) s'appuient sur la tâche des Strange Stories qui mobilise fortement les compétences verbales, insinuant une variable confondante qui limite les conclusions qui peuvent être tirées de ces résultats. Deuxièmement, comme remarqué de façon générale par Matthews et al. (2018), les rôles respectifs des deux prédicteurs importants que sont le langage et la théorie de l'esprit ne peuvent être distingués par ces études, qui ne les ont pas comparés directement. L'étude de Lecce et al. (2019) incluait bien une mesure de la compréhension lexicale, mais comme variable contrôle (permettant notamment de contrôler dans une certaine mesure la limite précédente), plutôt que d'évaluer son importance en tant que telle. L'étude de Whyte et Nelson (2015) constitue une exception notable à cette limite puisqu'elle a évalué l'importance de ces deux facteurs dans un groupe de 69 enfants à de 5 à 12 ans, mais en utilisant une mesure générale de compréhension du langage figuré. Cette mesure incluait des métaphores, mais également des énoncés ironiques et des requêtes indirectes par exemple. Dans cette étude, le langage (lexique et morphosyntaxe)

comme la théorie de l'esprit (évalué avec le test RME) étaient distinctement associées à la compréhension pragmatique, avec une association entre théorie de l'esprit et compréhension pragmatique qui se maintenait lorsque la compréhension langagière était contrôlée (comme Lecce et al., 2019), mais les contributions relatives de ces deux facteurs n'étaient pas explicitées.

### 3. Développement des implicatures scalaires

#### 3.1. Une immaturité persistante dans la compréhension des implicatures scalaires

Le développement des implicatures scalaires a également fait l'objet d'une large littérature. L'étude développementale de Noveck (2001) a même été un déclencheur des travaux chez l'adulte (*chapitre 1*) en montrant une immaturité persistante des enfants face aux implicatures scalaires, en comparaison de leur compréhension littérale, ce qui constitue un indicateur important de l'effort associé aux implicatures scalaires. Dans cette étude, Noveck (2001, Expérience 3) a proposé une tâche de vérification de phrases, en double aveugle, incluant des phrases sous-informatives (« quelques girafes ont un long cou ») et des phrases contrôles vraies (« toutes les girafes ont un long cou », « quelques oiseaux vivent en cage ») ou fausses (« tous les chiens ont des taches »). Des enfants de 8 ans, 10 ans et des adultes se sont prêtés à cette tâche. Alors que toutes les phrases contrôles étaient correctement jugées par l'ensemble des participants, les adultes étaient équivoques vis-à-vis des phrases sous-informatives (effet depuis considéré comme classique, voir le *chapitre 1*), avec 59% de rejets. Les enfants de 8 comme de 10 ans, en revanche, acceptaient largement ces phrases (dans 89 et 85% des cas, respectivement). Ils se montraient ainsi compétents avec le sens sémantique des quantifieurs, évalué dans les conditions contrôles, mais s'en tenaient à une interprétation sémantique de 'quelques' ('quelques-et-peut-être toutes') dans la condition expérimentale, à la différence des adultes qui étaient plus nombreux à manifester une interprétation pragmatique ('quelques-mais-pas-tous').

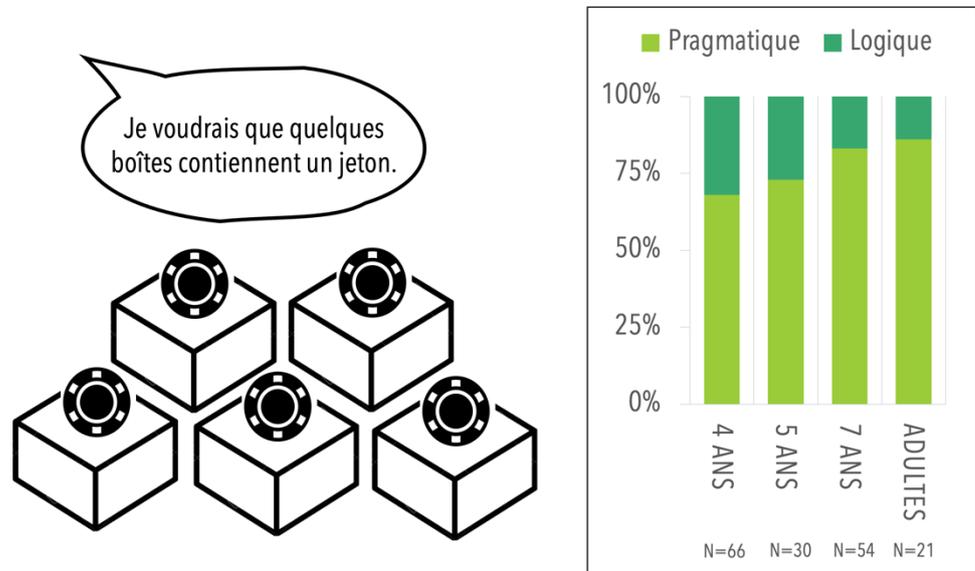
Cet effet a depuis été répliqué dans de nombreuses études (e.g. Barner et al., 2011; Chierchia et al., 2004; Guasti et al., 2005; Katsos & Bishop, 2011; Papafragou & Musolino, 2003; Pouscoulous et al., 2007). L'étude de Katsos et al. (2016) constitue probablement la plus robuste de ces réplifications, puisqu'elle incluait 1304 participants, locuteurs de 31 langues différentes, et confirmait de façon translinguistique cet effet développemental ; dans cette étude, les adultes rejetaient les énoncés sous-informatifs dans 84% des cas, contre seulement 51% des enfants (âgés de 5 ans), qui évaluaient pourtant correctement les phrases contrôles dans 82% des cas.

Par ailleurs, si l'étude princeps de Noveck (2001) demandait l'évaluation de phrases d'après des connaissances encyclopédiques (des connaissances sur les girafes, dans l'exemple ci-dessus), ce paradigme a été adapté pour demander une vérification à partir d'un contexte visuel, ce qui permet une manipulation expérimentale de ce

contexte, et pas seulement des termes utilisés. Papafragou et Musolino (2003) ont par exemple proposé à des participants d'évaluer des phrases telles que « quelques chevaux ont sauté par-dessus la barrière » après avoir fait la démonstration d'une scène dans laquelle *tous* les chevaux avaient sauté par-dessus une barrière (en animant des jouets). Alors que la quasi-totalité des adultes rejetaient les énoncés sous-informatifs, la majorité des enfants de 5 ans les acceptaient. Cette étude a inauguré une série de travaux visant à montrer que les implicatures scalaires pouvaient être facilitées chez l'enfant. Dans une deuxième expérience, Papafragou et Musolino (2003) ont en effet explicitement entraîné des enfants de 5 à 6 ans à détecter des énoncés inappropriés ou suboptimaux, par exemple en demandant si des phrases telles que « *c'est un animal avec quatre pattes* » pour désigner un chien était une réponse « correcte » ou « qui pourrait être meilleure », et en corrigeant les enfants avec la réponse plus informative en cas d'erreur (« *c'est un chien* »). Cet entraînement a nettement augmenté le taux de réponses pragmatiques des enfants (53% de rejets d'expressions en 'quelques' sous-informatives, contre seulement 13% à l'expérience 1), sans toutefois leur permettre d'atteindre des performances comparables à celles des adultes. Guasti et collaborateurs (2005) ont également entraîné de façon similaire des enfants plus âgés, de 7 ans, leur permettant cette fois de produire des réponses comparables à celles d'adultes. Toutefois, une re-visite une semaine après l'entraînement montrait que son effet s'était dissipé, avec des performances de nouveau comparables aux performances initiales.

D'autres manipulations expérimentales qu'un entraînement ont pu faciliter l'accès aux implicatures scalaires des enfants. Foppolo et al. (2012) ont notamment montré que rendre l'alternative à 'quelques' plus accessible augmentait la probabilité qu'un enfant dérive une implicature scalaire. Ainsi, demander de vérifier une phrase vraie avec le terme 'tous', puis une phrase fautive avec le terme 'tous' (pour laquelle 'quelques' était approprié), avant une phrase sous-informative en 'quelques' (pour laquelle 'tous' était approprié) augmente le rejet de ces phrases de 42% à 70% chez des enfants de 5 ans (contre la quasi-totalité des adultes). Pouscoulous et al. (2007) ont également proposé différentes modifications visant à faciliter l'accès à ces implicatures, en limitant la charge cognitive induite par le paradigme expérimental. Les tâches de jugement de phrases classiques demandent en effet de porter un jugement métalinguistique sur une production langagière, ce qui même pour des adultes n'est pas trivial (*chapitre 1*), et qui pourrait expliquer les difficultés persistantes des enfants avec cette tâche. Pouscoulous et al. (2007, expérience 2) ont ainsi modifié le paradigme expérimental original afin de recueillir une réponse non sous la forme d'un jugement, mais d'une action. Par exemple, si 5 boîtes sont présentées à un enfant, dont 2 contiennent un jeton et qu'on demande à cet enfant « *Je voudrais que toutes les boîtes contiennent un jeton* », celui-ci pourrait ajouter des jetons dans les trois boîtes vides, ce qui permettrait de déterminer que l'enfant considère que la situation initiale n'est pas compatible avec la consigne. Autour de cette situation, Pouscoulous et al. font varier la disposition initiale des jetons ainsi que le quantifieur utilisé, de sorte qu'il existe aussi des conditions où aucune action n'est nécessaire (par exemple, la même consigne est formulée alors que 5/5 boîtes ont déjà un jeton). La situation critique est celle où toutes

les boîtes contiennent initialement un jeton, mais où on demande à l'enfant « *Je voudrais que quelques boîtes contiennent un jeton* » (figure 2-4, gauche). Si le participant considère que 'quelques' est compatible avec 'tous' (interprétation logique), alors aucune action n'est nécessaire, mais si 'quelques' est interprété pragmatiquement ('quelques-mais-pas-tous'), alors il est nécessaire de retirer au moins un jeton.



**Figure 2-4 :** Illustration de l'item critique du paradigme basé sur l'action proposé par Pouscoulous et al. (2007, expérience 2, gauche) et proportion de chaque groupe répondant de façon pragmatique (droite) (données extraites de la publication originale)

Avec ce paradigme, l'ensemble des conditions contrôles étaient largement réussies, y compris chez les enfants les plus jeunes qui étaient évalués (4 ans). Dans la condition critique (figure 2-4), 86% des adultes montrent une interprétation pragmatique, contre 83% des enfants de 7 ans, 73% des enfants de 5 ans et 68% des enfants de 4 ans. Cela contraste donc fortement avec les performances classiquement observées, dans lesquelles les enfants répondent de façon nettement moins pragmatique, y compris à un âge plus élevé. Pouscoulous et al. (2007) avaient notamment répliqué cet effet classique dans une première expérience, avec des énoncés similaires, où seuls 11% des enfants de 9 ans répondaient de façon pragmatique (contre 47% des adultes). Outre la modification du format de réponse, ce nouveau paradigme expérimental réduisait aussi la charge cognitive limitant les distracteurs et en employant le quantifieur 'quelques', plus fréquent et acquis plus précocement que 'certaines', un autre équivalent acceptable de l'anglais 'some'. Une troisième expérience, enfin, confirmait qu'utiliser 'quelques' plutôt que 'certaines' facilitait l'inférence pragmatique. Cette étude démontre ainsi clairement que différentes manipulations limitant le coût cognitif de la tâche expérimentale facilite l'accès aux implicatures scalaires d'enfants même jeunes. Toutefois, il apparaît également clairement que, si cet accès peut être facilité, l'effet développemental reste bien présent dans les données, avec une augmentation de l'accès aux implicatures avec l'âge (figure 2-4, droite). Similairement, Miller et al.

(2005) avaient également montré avec une tâche de dessin qu'un paradigme basé sur l'action facilitait l'accès des enfants aux implicatures scalaires, tout en préservant un effet développemental clair. Dans cette étude, trois visages dépourvus de bouche étaient par exemple présentés aux participants, tandis qu'il leur était demandé « *[d'utiliser] le crayon et de [rendre] quelques visages contents* ». Dans cette condition, la moitié des productions d'enfants de 4 à 5 ans étaient considérées comme pragmatiques (un visage n'était pas dessiné sur *tous* les visages), contre la quasi-totalité de celles des adultes.

Une autre modalité d'évaluation évitant toute demande métalinguistique est le paradigme d'eye-tracking évoqué au [premier chapitre](#), que ses auteurs ont également appliqué à une population d'enfants de 5-6 ans (Y. T. Huang & Snedeker, 2009b). Ce paradigme, qui révélait que des adultes utilisent *en direct* un quantifieur scalaire au sein d'une phrase telle que '*Montre la fille qui a quelques chaussettes*' pour identifier le référent, sans attendre la fin de la phrase, a fait apparaître un effet développemental important. En effet, une telle anticipation n'a pas pu être observée chez les enfants. Ceux-ci ne regardaient préférentiellement la cible (l'image où la fille a 2/3 chaussettes), qu'à partir du moment où la phrase n'était plus ambiguë. En d'autres termes, après avoir entendu '*Montre la fille qui a quelques...*', l'image où une fille a *tous* les ballons demeurait une cible potentielle pour les enfants, à la différence des adultes. Avec cette mesure implicite et ce dispositif peu coûteux cognitivement, les enfants ne montraient donc pas de signes compréhension d'implicatures scalaires en direct, lors de la présentation orale d'une phrase. D'un point de vue méthodologique, on pourrait opposer que le temps disponible pour manifester une inférence pragmatique en direct est limité, dans cette tâche, jusqu'au moment où le phonème désambiguïseur est énoncé (environ 670 ms); des signes d'interprétations pragmatiques auraient pu émerger si les enfants avaient disposé de puis de temps (Katsos & Bishop, 2011, voir [section suivante](#)). Cependant, les enfants se montraient aussi rapides que les adultes pour interpréter en direct les quantifieurs lorsque leur sens sémantique était suffisant. Ainsi, ces données seraient expliquées ou bien par une absence d'inférence pragmatique en direct, ou bien par un retard spécifique et plus important que les adultes, ce qui dans tous les cas est en faveur d'une immaturité persistante vis-à-vis de ces phénomènes.

Ainsi, bien que modulé par différents facteurs, l'effet développemental observé dans la propension à réaliser des implicatures scalaires demeure confirmé par une large série d'études. Ce retard est classiquement interprété comme un délai dans la mise en place des opérations pragmatiques, une immaturité pragmatique persistante, dont de multiples causes ont été postulées. Certains postulent que les enfants ne disposeraient pas de suffisamment de ressources cognitives générales pour mettre en place tout ou partie des étapes inférentielles (Pouscoulous et al., 2007; Reinhart, 2004), comme le suggère certains des travaux présenté ci-dessus. D'autres proposent que ce retard soit expliqué par une maîtrise tardive des concepts sémantiques et lexicaux des termes utilisés (Barner et al., 2011; Guasti et al., 2005). Alternativement, des résultats

suggèrent que les difficultés à réaliser des implicatures scalaires soient associées à une immaturité de la théorie de l'esprit, plutôt que des représentations sémantiques, puisque ces inférences nécessitent de comprendre (au moins implicitement) qu'un locuteur fournit autant d'information qu'il possède, et donc de se représenter ses croyances (Papafragou et al., 2018). À l'inverse de ces approches, et similairement au mouvement observé dans le domaine des métaphores (voir [section 2.1](#)), l'hypothèse de la tolérance pragmatique (Katsos, 2014; Katsos & Bishop, 2011) postule que les enfants possèderaient précocement toutes les compétences pragmatiques nécessaires à la dérivation d'implicatures scalaires, mais que celles-ci seraient masquées par les paradigmes expérimentaux qui sont classiquement utilisés.

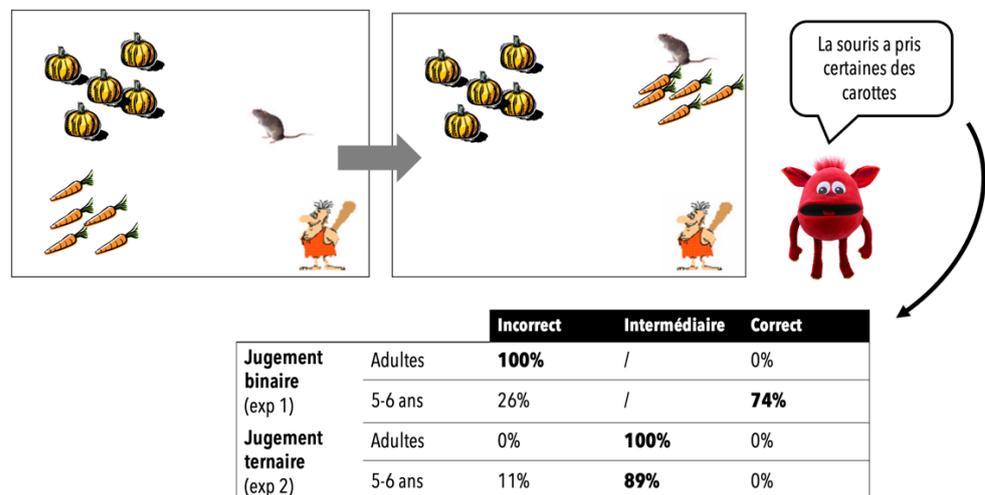
### 3.2. Hypothèse de la tolérance pragmatique

L'hypothèse de la tolérance pragmatique s'appuie sur une critique méthodologique des paradigmes de jugement de phrases classiquement utilisés pour évaluer si un individu comprend ou non des implicatures scalaires. Selon Katsos & Bishop (2011), ces tâches ne devraient pas être directement assimilées à une mesure de l'inférence scalaire. En effet, un énoncé sous-informatif tel que « *Certains éléphants ont une trompe* » est à la fois logiquement vrai et sous-optimal en termes d'informativité. Le fait de devoir juger si un tel énoncé est correct ou non oblige donc un participant à sélectionner une stratégie, selon qu'il se base sur la vérité propositionnelle, ou sur l'optimalité informative, pour émettre son jugement. Un participant qui ne réalise pas d'implicature scalaire et ne perçoit pas la sous-informativité d'un énoncé le jugerait donc correct, mais un participant qui perçoit cela pourrait porter le même jugement, s'il décide de se focaliser sur sa vérité propositionnelle plutôt que sur son informativité. À l'inverse, rejeter un énoncé sous-informatif pourrait être la conséquence de la dérivation d'une implicature scalaire, mais aussi, plus simplement, de la conscience de sa sous-informativité. C'est pourquoi Katsos et Bishop affirment qu'une tâche de jugement de phrases binaire ne permet pas de mesurer *directement* comment des participants interprètent des termes scalaires, ni s'ils sont sensibles à l'informativité d'un énoncé. Cette critique a par ailleurs très récemment trouvé un soutien empirique via des expérimentations chez l'adulte. Kissine et De Brabanter (2023) ont en effet montré qu'immédiatement après avoir jugé un énoncé sous-informatif faux (par exemple « *Le joueur a marqué quelques points* »), des adultes peuvent pourtant sélectionner comme appropriée une paraphrase de cet énoncé explicitant une interprétation logique (« *Le joueur a marqué quelques points, et peut-être tous* ») plutôt qu'explicitant l'implicature scalaire (« *Le joueur a marqué quelques points, mais pas tous* »).

En s'appuyant sur cette critique méthodologique, Katsos & Bishop (2011) introduisent la possibilité théorique que, dans des tâches de jugements classiques, les jeunes enfants perçoivent qu'un énoncé est sous-informatif, mais qu'ils ne considèrent pas cela comme une raison suffisante de les rejeter (à la différence des adultes). En d'autres termes, ils possèderaient la compétence pragmatique nécessaire à la détection de la sous-informativité et à la compréhension d'implicatures scalaires, mais montreraient

une certaine *tolérance pragmatique* face à des violations de cet ordre. Cette hypothèse expliquerait ainsi le fossé apparent entre les prérequis et les habiletés pragmatiques précoces des enfants décrits à la [section 1.1](#), et les performances classiquement observées dans les tâches d'implicatures scalaires.

Pour soutenir cette proposition, Katsos & Bishop (2011) répliquent dans un premier temps le résultat classiquement observé, avec une tâche dans laquelle une poupée, décrite comme apprenant l'anglais, demande au participant d'évaluer ses productions (en anglais) pour qu'elle s'améliore. La poupée décrit des actions réalisées par des personnages sur un écran d'ordinateur, avec des énoncés vrais, faux ou sous-informatifs tels que « *La souris a pris certaines des carottes* » quand une souris a pris l'ensemble des carottes disponibles (figure 2-5, en haut). L'examineur demande ensuite au participant si la poupée a raison ou non. Dans ces conditions, la totalité des adultes rejettent les énoncés sous-informatifs, tandis que seuls 26% des enfants de 5-6 ans le font, répliquant donc clairement l'effet développemental classique (tandis que toutes les conditions contrôles sont bien réussies par les deux groupes).



**Figure 2-5 :** Représentation du paradigme utilisé par Katsos et Bishop (2011, en haut) et des résultats obtenus par les groupes aux items critiques en fonction de la modalité de réponse (binaire ou ternaire, en bas; données extraites de la publication originale pour la condition scalaire uniquement)

Mais Katsos et Bishop introduisent ensuite une modification critique dans le paradigme classique de jugement de phrase, qui permet cette fois d'évaluer directement la sensibilité des participants à l'informativité d'une phrase, et donc à l'ambiguïté d'énoncés sous-informatifs. Plutôt que de déterminer si la phrase de la poupée est *correcte* ou *non* (jugement binaire), l'expérimentateur demande au participant de la récompenser avec des fraises (présentées comme fruit préféré de la poupée) de différentes tailles. Trois fraises sont présentées, chacune étant 2 fois plus large que la précédente, comme une 'petite fraise, une 'grosse fraise' et une 'énorme fraise'. Cette échelle de jugement ternaire permet donc de distinguer des énoncés qui seraient plus corrects que des énoncés faux, mais moins que des énoncés optimalement informatifs. Avec ce nouveau format de réponse, Katsos et Bishop observent que les

enfants comme les adultes sélectionnent la petite fraise pour les énoncés faux et l'énorme fraise pour les énoncés corrects et optimaux, tandis que les deux groupes favorisent la réponse intermédiaire pour les énoncés sous-informatifs (dans 100% des cas chez les adultes et 89% chez les enfants, cf. figure 2-5, en bas).

Katsos et Bishop démontrent ainsi que des enfants de 5 à 6 ans peuvent être capable, dans de telles conditions, de montrer des performances similaires aux adultes, manifestant ainsi qu'ils peuvent être aussi sensibles à l'informativité d'un énoncé, mais que cette compétence n'est pas visible dans les paradigmes de jugement binaires. Katsos et Bishop proposent ainsi que dans les tâches classiques, les enfants fassent preuve d'une tolérance aux violations pragmatiques des énoncés, tout en percevant ces violations. Cette hypothèse expliquerait les effets positifs des entraînements décrits ci-dessus (Guasti et al., 2005; Papafragou & Musolino, 2003), qui exploiteraient des compétences pragmatiques présentes (et qu'un bref entraînement ne pourrait pas suffire à faire émerger) en incitant les enfants à rejeter des violations pragmatiques qu'ils seraient bien capables de percevoir.

## Développement typique et atypique

Tandis que les travaux présentés ci-dessus s'attachent à décrire et expliquer le développement typique des inférences pragmatiques, une propension non négligeable d'enfants présente un neurodéveloppement atypique de différentes fonctions cognitives, pouvant impacter leur compréhension pragmatique (Cummings, 2016, 2021). C'est le cas de l'autisme (e.g., Happé, 1993), du trouble développemental du langage (e.g., Norbury, 2005), du trouble du déficit de l'attention avec ou sans hyperactivité (e.g., Caillies et al., 2014) ou encore des troubles des apprentissages (e.g., Cappelli et al., 2018). Les troubles du neurodéveloppement concernent au minimum 5 % de la population (voire bien plus selon certaines estimations, voir Francés et al., 2022), or les inférences pragmatiques ont une importance fonctionnelle considérable (cf. *chapitre 1*), ce qui fait de cette question un enjeu de santé publique important. Mieux décrire les mécanismes pouvant entraver ces compétences est donc essentiel pour mieux identifier, prévenir et réduire l'impact de ces importantes sources de handicap social et scolaire puis professionnel.

Par ailleurs, si l'étude des différences interindividuelles peut éclairer notre compréhension des inférences pragmatiques, comme nous l'avons vu dans ce chapitre pour le développement typique, cette approche bénéficie également de la comparaison de personnes de différentes catégories diagnostiques. Certains travaux se sont notamment portés sur des conditions acquises ou apparaissant à l'âge adulte, telles que la schizophrénie (Bambini, Arcara, et al., 2016; Champagne-Lavau & Stip, 2010) ou les lésions cérébrales droites (Champagne-Lavau & Joannette, 2009). Dans ce travail, nous resterons focalisés sur le développement, en nous tournant dans le prochain

chapitre vers une condition neurodéveloppementale dans laquelle de nombreuses recherches ciblant les inférences pragmatiques ont été conduites : l'autisme.

## Chapitre 3 • Inférences pragmatiques et autisme

Elle m’a demandé de garder un œil sur ses affaires. Ça me fait toujours penser à retirer mon œil de ma tête pour le mettre sur les affaires, mais je ne lui ai pas dit que c’est à ça que je pensais. Je sais ce que cette expression signifie socialement, mais c’est une signification idiote. (Moon, 2002, pp. 153–154 [traduction personnelle])

Dans cet extrait du roman *Speed of Dark*, Elizabeth Moon donne la parole à un narrateur autiste (fictif), qui rapporte sa compréhension *au premier degré* d’une expression figurée, en l’occurrence d’une métonymie conventionnalisée (‘garder un œil’). Ce passage illustre bien la tendance typiquement associée à l’autisme de s’attacher au sens littéral du langage, plutôt qu’à l’implicite qu’il véhicule, que nous examinerons tout au long de ce chapitre. Après avoir évoqué comment l’autisme peut être défini de façon générale (*section 1.1*), nous nous focaliserons sur cette particularité du rapport au langage qui est intrinsèquement associée à nos représentations de l’autisme (*section 1.2*). Toutefois, nous verrons ensuite que la littérature expérimentale pousse à nuancer cette affirmation, en nous penchant plus particulièrement sur la compréhension des métaphores dans un premier temps (*partie 2*), puis des implicatures scalaires (*partie 3*). Nous verrons ensuite en quoi ces données questionnent également nos modèles pragmatiques plus généralement (*partie 4*). Pour désigner les personnes concernées, nous utiliserons l’expression de *personnes autistes*, et détaillons les raisons de ce choix non anecdotique dans *l’encart 3*.

### **Encart 3 :** Comment désigner les personnes avec un diagnostic d’autisme ?

Les termes avec lesquels nous désignons collectivement les personnes concernées par l’autisme ont des conséquences dans le monde réel, au-delà de simples considérations linguistiques (Botha et al., 2021; Vivanti, 2020). Ils influencent les représentations collectives et, ce faisant, les attitudes stigmatisantes voire violentes de la société envers les personnes autistes, mais aussi les politiques publiques, les pratiques cliniques et les orientations de la recherche. Cette question est l’objet de débats au sein de la communauté scientifique, qui opposent deux approches principales visant toutes deux à limiter les stéréotypes et les violences dont les personnes concernées peuvent être victimes. L’approche focalisée sur la personne propose de parler de ‘personne avec autisme’ (*‘person with autism’, person-first language*), tandis que l’approche focalisée sur l’identité pousse plutôt à parler de ‘personne autiste’ (*‘autistic person’, identity-first language*). L’approche focalisée sur la personne s’est développée à partir des années 1970 pour mettre en avant la personne plutôt que son trouble, avec ses expériences et son unicité, en réponse à des discours déshumanisants, essentialisants et violents. Cette approche continue de trouver des partisans (e.g. Vivanti, 2020) qui y voient un moyen d’empouvoirement des personnes concernées. D’autres, en revanche, opposent que cette approche sépare linguistiquement la personne de l’autisme, véhiculant la possibilité qu’une personne puisse être séparée de son autisme dans le monde réel (ce qui est impossible). De plus, cela induirait l’idée que cette séparation serait souhaitable,

l'autisme constituant un objet négatif, à combattre. Cette approche autoriserait ainsi des formules telles que 'vaincre l'autisme', perçue très violemment par les personnes concernées, véhiculant des stéréotypes négatifs et socialement néfastes de l'autisme. L'approche focalisée sur la personne est ainsi critiquée par le mouvement de la neurodiversité qui cherche à dépathologiser l'autisme (sans nier le handicap qu'il peut engendrer, cf. Chapman, 2019), et qui oppose une approche focalisée sur l'identité (voir figure 3-1). Selon cette approche, l'autisme fait partie intégrante de l'identité de la personne, elle ne peut être annulée et doit donc être acceptée et respectée. S'il n'y a donc pas de consensus dans la communauté scientifique, la communauté des personnes concernées utilise elle-même de multiples termes. Dans plusieurs études anglo-saxonnes, l'approche focalisée sur la personne apparaît toutefois comme la moins préférée et celle jugée la plus offensante par les personnes avec un diagnostic d'autisme (Bury et al., 2020; Kenny et al., 2016). Une récente étude confirme cette préférence dans la communauté francophone (Geelhand et al., 2023). Il faut toutefois noter que ces choix apparaissent différents dans d'autres communauté linguistiques/culturelles, et qu'ils sont influencés par différentes caractéristiques des répondants telles que l'âge, le QI ou le caractère prononcé de l'autisme (Buijsman et al., 2023). Les participants qui expriment leurs préférences dans ces études ne sont généralement pas représentatifs de l'ensemble des personnes autistes, n'incluant typiquement pas les personnes peu verbales ou oralisantes, et de façon générale celles ayant un degré de handicap plus important. Or, certains opposent qu'adopter l'approche focalisée sur l'identité, à la demande d'un sous-groupe, pourrait n'être profitable qu'à ce sous-groupe et nuire aux sous-groupes qui ne se sont pas exprimés (Vivanti, 2020). En l'absence de consensus et tout en reconnaissant la légitimité de différentes approches, nous suivons dans cette thèse le raisonnement de Botha et al. (2021) et adoptons une terminologie focalisée sur l'identité, et ce d'autant plus que ce travail de thèse porte principalement sur le sous-groupe qui a pu exprimer ses préférences dans les études citées ci-dessus.



**Figure 3-1 :** Illustration des perceptions induites par différents choix terminologiques pour évoquer les personnes ayant un diagnostic d'autisme, traduite d'un tweet d'@Autistic\_Cherry

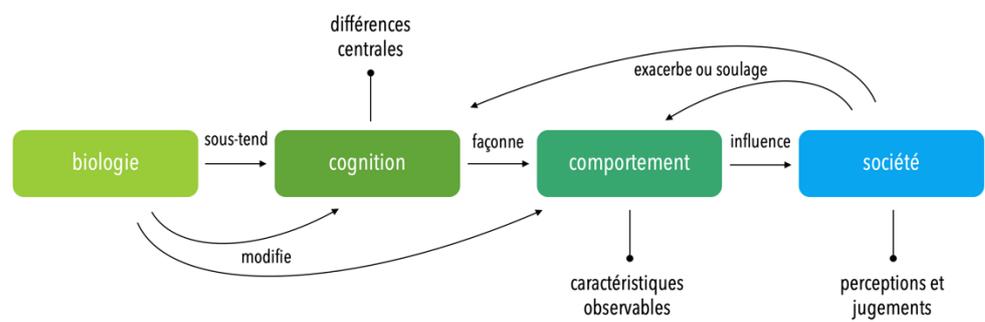
## 1. Le second degré : au cœur de l'autisme ?

### 1.1. Définir l'autisme

L'autisme est une condition complexe et hétérogène, dont la définition est un défi en tant que tel qui a fait l'objet de nombreux travaux (voir Fletcher-Watson & Happé, 2019). L'autisme peut être décrit au niveau biologique, au niveau cognitif (incluant ici tout le fonctionnement de l'esprit, sans exclure les aspects émotionnels par rapport aux

aspects rationnels) ou au niveau comportemental, chaque niveau sous-tendant le niveau suivant (figure 3-2). En plus de ces niveaux, les modèles sociaux du handicap poussent à prendre également en compte le niveau sociétal, déterminé par l'ensemble des comportements des individus d'une population, mais ayant une rétroaction sur les niveaux précédents. Si les niveaux comportementaux et biologiques sont directement observables, et donc à privilégier pour établir une définition fiable et consensuelle, ce n'est pas le cas du niveau cognitif, qui fait pourtant le lien entre fonctionnement biologique et comportement. De plus, de nombreuses théories cognitives de l'autisme existent, sans consensus clair, et ne peuvent donc pas être utilisées comme terrain commun pour définir cet objet d'étude. C'est toutefois le niveau qui nous intéressera principalement dans cette thèse, puisque nous travaillons sur les mécanismes *cognitifs* des inférences pragmatiques.

Au niveau biologique, la forte implication d'une composante génétique dans l'autisme ne fait plus débat (e.g. Tick et al., 2016), et de nombreuses atypies ont également été identifiées au niveau neuroanatomique et neurofonctionnel (Amaral et al., 2008; Geschwind, 2008; O'Reilly et al., 2017). Cependant, aucun marqueur suffisamment sensible et spécifique n'a pu être identifié à ce niveau, avec une forte hétérogénéité à l'intérieur de ce groupe. S'il est établi que l'autisme a des fondations génétiques importantes, menant à des différences neurologiques, la description la plus consensuelle de l'autisme est donc opérée au niveau comportemental. Les observations du comportement sont ainsi utilisées à des fins diagnostiques dans les deux classifications nosologiques de références, le DSM-5-TR de l'American Psychiatric Association (2022) et la 11<sup>ème</sup> Classification Internationale des Maladies (CIM-11) de l'Organisation Mondiale de la Santé (World Health Organization, 2019).



**Figure 3-2 :** Différents niveaux d'explication du neurodéveloppement, avec les trois niveaux classiquement considérés pour caractériser l'autisme (en vert), et le niveau sociétal proposé par les modèles sociologiques (en bleu) (figure adaptée de Fletcher-Watson & Happé, 2019, reproduction autorisée par Taylor and Francis Group, LLC, une filiale de Informa plc)

Ces deux classifications, largement concordantes, définissent le trouble du spectre de l'autisme (TSA, ou simplement *autisme* dans le reste de ce travail<sup>11</sup>) par l'association

<sup>11</sup> Le terme de *Trouble du Spectre de l'Autisme* est utilisé afin de marquer l'hétérogénéité des présentations de cette condition (Wing, 1997), à la fois entre les individus et, pour un même individu,

de caractéristiques dans deux domaines distincts : des atypies dans la communication, l'interaction et les relations sociales d'une part, et des comportements, intérêts et activités inhabituellement restreints et répétitifs, et/ou des particularités sensorielles d'autre part. L'autisme appartient au groupe des troubles du neurodéveloppement, caractérisés par des particularités persistantes, présentes précocement dans le développement de la personne, et à l'origine de limitations dans son fonctionnement personnel, social, professionnel et/ou scolaire. Ces classifications demandent en outre de spécifier si l'autisme co-occure avec deux autres conditions très fréquemment associées : le trouble du développement intellectuel et le trouble du langage. Des travaux montrent par exemple un recouvrement partiel de particularités langagières entre le trouble du spectre de l'autisme et des troubles du langage (e.g. Durrleman & Delage, 2016). L'ensemble de ces signes peuvent se combiner de façons multiples et variables, faisant de l'autisme une condition particulièrement hétérogène, ce qui complexifie fortement son étude. De ce fait, la littérature scientifique sur les inférences pragmatiques s'est quasi-exclusivement focalisée sur le sous-groupe des personnes sans handicap intellectuel et sans trouble du langage (auparavant désignées par les expressions d'autisme de haut niveau et du syndrome d'Asperger). Ce sera donc également le cas de cette thèse.

Les tableaux cliniques des personnes concernées sont d'autant plus complexes que l'autisme est aussi fréquemment associé à d'autres troubles du neurodéveloppement comme les troubles du déficit de l'attention / hyperactivité (TDAH), les troubles des apprentissages ou les troubles du développement des coordinations (Levy et al., 2010; Schieve et al., 2012), ainsi qu'à des troubles psychiatriques comme la dépression (Skokauskas & Gallagher, 2010) ou l'anxiété (Van Steensel et al., 2011), et à des vulnérabilités somatiques notamment au niveau du sommeil, de l'alimentation ou du système digestif (Hirata et al., 2016; McElhanon et al., 2014). La prévalence de l'autisme en fait un enjeu important, puisqu'elle est actuellement estimée aux alentours de 1% (malgré des variations considérables entre les études, voir Baxter et al., 2015; Elsabbagh et al., 2012).

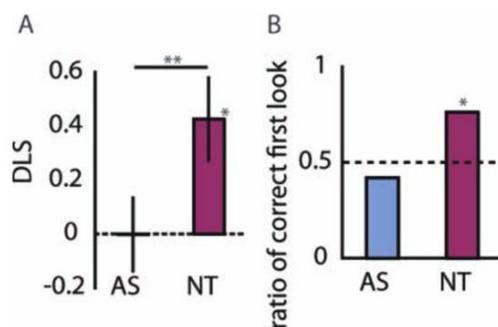
Les modèles cognitifs de l'autisme tentent de lier l'hétérogénéité observée au niveau comportemental et au niveau biologique en un modèle plus global. Parmi de nombreuses propositions, il sera utile d'évoquer ici le modèle du *déficit en théorie de l'esprit*, théorie dite de la « cécité mentale », qui a eu un impact particulier. Dans une étude fondatrice, Baron-Cohen et al. (1985) ont proposé à des enfants autistes et à plusieurs groupes contrôles une tâche standard de fausses croyances de premier ordre pour évaluer leur théorie de l'esprit. Alors que les enfants autistes échouaient majoritairement à cette tâche, un groupe contrôle d'enfants avec une déficience

---

en fonction du développement et du contexte. Cette appellation marque également le regroupement de catégories diagnostiques auparavant distinctes comme l'autisme typique ou atypique, le syndrome d'Asperger, le trouble envahissant du développement non-spécifié, etc. Par ailleurs le terme de *trouble* est décrié par la communauté des personnes concernées, aussi utiliserons nous le terme plus simple d'*autisme* ou celui plus générique de *condition*.

intellectuelle d'un âge similaire réussissait majoritairement la tâche, de même qu'un groupe d'enfants à développement typique beaucoup plus jeunes. Sur cette base, les auteurs ont proposé qu'un déficit méta-représentationnel en théorie de l'esprit soit la caractéristique essentielle et primaire de l'autisme (voir aussi Baron-Cohen, 1997). Cette théorie a eu un impact considérable puisqu'elle proposait un cadre théorique permettant des prédictions expérimentales claires, exploitables notamment dans les théories de la communication, comme nous le verrons plus tard ; de plus, elle offrait une cohérence à des manifestations comportementales jusqu'alors éclatées, étant ainsi facilement transférable dans les pratiques cliniques et éducatives. Cette hypothèse a depuis été fortement nuancée, et n'est plus considérée comme un modèle *explicatif* de l'autisme puisqu'un tel déficit n'est ni *universel* dans cette population, ni *spécifique* à cette condition (de telles difficultés sont par exemple fréquentes dans d'autres troubles du neurodéveloppement comme les troubles du langage, des conditions sensorielles comme la déficience auditive, ou psychiatriques comme la schizophrénie), et qu'on manque par ailleurs d'arguments pour affirmer le caractère *primaire* de ces déficits sur l'ensemble des manifestations de l'autisme. En outre, les modèles ont évolué d'hypothèses impliquant un ou des *déficits*, vers des modèles proposant des *différences* et intégrant leur caractère développemental (Fletcher-Watson & Happé, 2019; Georgieff, 2017).

L'autisme reste toutefois associé à des différences (plutôt qu'un déficit ou qu'une absence) dans les processus de représentation des états mentaux, mais dont l'observation peut nécessiter des paradigmes plus subtils que les tâches standards de fausses croyances (e.g. Baltazar et al., 2021). Par exemple, nous avons évoqué au [chapitre 2](#) des mesures *implicites* de théorie de l'esprit, telle que des mesures d'eye-tracking associées à des situations de type fausses croyances. Celles-ci ont mis en évidence que des enfants à développement typique utilisaient spontanément les représentations d'autrui pour prédire leur comportement, bien avant de pouvoir répondre à une question explicite sur les croyances d'un personnage (24 mois, Southgate et al., 2007). Senju et al. (2009) ont proposé ce paradigme à des adultes autistes, sans handicap intellectuel et sans trouble du langage, et à un groupe contrôle non-autiste. Alors que ces participants ont correctement répondu à des tâches explicites de fausses croyances, dans cette tâche implicite les participants autistes ne regardaient pas plus la cible que le distracteur (figure 3-3, en bleu), à la différence du groupe contrôle. Ils ne manifestaient donc pas d'utilisation *spontanée* de la théorie de l'esprit pour anticiper le comportement du personnage. Senju (2012) propose donc que cette compétence socio-cognitive puisse être fonctionnelle dans l'autisme, mais peu spontanément recrutée dans le traitement des informations sociales.



**Figure 3-3 :** Données d'eye-tracking d'adultes neurotypiques (NT) et autistes (AS) dans une tâche de fausses croyances implicites, représentant le rapport entre le temps de fixation de la cible et celui du distracteur (A) et la proportion de premiers regards vers la cible plutôt que le distracteur (B) (lignes pointillées = niveau de chance ; figure extraite de Senju, 2012)

Ces différences occupent encore une place importante dans les modèles plus récents de l'autisme, dont deux seront évoqués ici. La théorie du cerveau mâle extrême (Extreme Male Brain theory, EMB) de Baron-Cohen (2010) postule que le fonctionnement cérébral et cognitif humain est constitué de deux grands systèmes parallèles, l'un tourné vers la recherche de patterns systématiques dans l'environnement (systématisation) et l'autre vers la compréhension d'autrui (empathie). Ces deux systèmes s'équilibreraient, avec toutefois une dominance relative de la systématisation chez les hommes et de l'empathie chez les femmes. L'autisme serait alors caractérisé par une systématisation extrême, au détriment du système empathique, lui conférant ses particularités, avec ses faiblesses mais également ses forces cognitives et comportementales. Alternativement, la théorie de la motivation sociale (Chevallier et al., 2012) propose que, plutôt qu'un déficit dans les compétences sociales en elles-mêmes, l'autisme soit caractérisé par un défaut de motivation sociale qui limite la propension avec laquelle ces individus se tournent vers les stimuli sociaux, et mobilisent spontanément leurs compétences pour les traiter.

D'autres considérations ont en outre nuancé ces approches. Notamment le fait que dans l'immense majorité de ces études, des personnes autistes se prêtent à des recherches conduites par des chercheurs et expérimentateurs non-autistes. Or, le fait d'appartenir à un groupe social différent de leur interlocuteur (pour les participants autistes mais pas pour les participants contrôles neurotypiques) peut induire des modifications comportementales en lui-même. Ce contraste entre un groupe interne (*in-group*) auquel un individu appartient et un groupe externe (*out-group*) auquel il n'appartient pas a notamment été observé sur des comportements classiquement considérés comme atypiques dans l'autisme, comme l'imitation. Par exemple, Buttelmann et al. (2013) ont expérimentalement créé cet effet auprès d'enfants à développement typique de 14 mois, en les exposant à un locuteur leur racontant une histoire dans leur langue maternelle (se positionnant dans un groupe interne) ou dans une langue étrangère (se positionnant dans un groupe externe). Les enfants imitaient près de deux fois plus fréquemment une action réalisée par l'expérimentateur lorsqu'il se positionnait dans un groupe interne que dans un groupe externe. Il est donc possible qu'une partie des

effets observés expérimentalement dans des études comparant des participants autistes et non-autistes puissent être attribués à ce contraste entre groupe interne et externe. Dans la lignée de ces idées, Milton (2012) pose le *problème de la double empathie* en rappelant que les difficultés d'empathie attribuées aux personnes autistes sont typiquement observées à partir d'interactions ou de stimuli impliquant des personnes non-autistes. Mais qu'à l'inverse, les personnes non-autistes manifestent aussi peu d'empathie envers les personnes autistes. Des adultes non-autistes peuvent par exemple être en difficultés pour attribuer des états mentaux à des personnes autistes, par rapport à des personnes neurotypiques (Edey et al., 2016). Le problème de la double empathie propose ainsi une vision des difficultés d'interactions entre autistes et non-autistes comme un problème de compréhension mutuelle et de réciprocité, plutôt qu'une responsabilité et une caractéristique intrinsèque des personnes autistes (voir Crompton et al., 2020). Ces approches illustrent ainsi l'importance de considérer également le niveau sociétal représenté sur la figure 3-2 lorsqu'on s'interroge sur la nature de l'autisme.

L'autisme est donc un objet d'étude complexe, dont les descriptions peuvent porter sur différents niveaux d'explication et ont largement évolué dans le temps. Dans ce contexte, nous proposons désormais de nous tourner vers la place particulière qu'a occupé la compréhension de l'implicite, et donc les inférences pragmatiques, dans la caractérisation de l'autisme, depuis ses premières descriptions modernes.

## **1.2. Une compréhension fondamentalement littérale ?**

Des premières descriptions de l'autisme par Kanner (1943) et Asperger (1944/1991) aux classifications actuelles (DSM-5-TR, APA, 2022) on retrouve, parmi les difficultés de communication et d'interaction, des difficultés pragmatiques, et plus particulièrement une interprétation excessivement littérale du langage. Pour décrire Donald, son tout premier cas, Kanner (1943) écrit ainsi : « *Les mots avaient pour lui une signification spécifiquement littérale, un sens inflexible* » (p. 219). Si les enfants qu'il décrivait étaient relativement peu verbaux, Asperger (1944/1991) à l'inverse évoque des enfants avec de bonnes compétences langagières, qui pouvaient étonnamment parler « *comme un adulte* » (p. 39), mais qui étaient en difficulté pour utiliser ces compétences dans la communication : « *le langage autistique n'est pas dirigé vers le destinataire mais est souvent prononcé comme dans un espace vide* » (p. 70). Il décrit également « *l'utilisation d'expressions nouvelles ou partiellement restructurées, qui peuvent souvent être particulièrement perspicaces mais aussi, bien sûr, particulièrement absconses* », et plus particulièrement sur le versant de la compréhension : « *Une autre caractéristique des enfants autistes est leur absence de sens de l'humour. Ils ne comprennent pas les blagues* » (p. 82). En d'autres termes, Asperger décrit des difficultés pragmatiques, en l'absence de difficultés langagières formelles. Ces caractéristiques sont précisément spécifiées dans le DSM-5-TR, qui décrit que « *de nombreux individus ont des troubles du langage, allant de l'absence complète de parole à des retards de langage, une mauvaise compréhension, un*

*discours écholalique ou UN LANGAGE GUINDE OU EXCESSIVEMENT LITERAL. Même lorsque les compétences langagières formelles (e.g. vocabulaire, grammaire) sont intactes, L'UTILISATION DU LANGAGE POUR LA COMMUNICATION SOCIALE RECIPROQUE EST ENTRAVEE dans le trouble du spectre de l'autisme »* (p. 61, traduction et emphases personnelles).

Cette caractéristique de l'autisme a ainsi infiltré les représentations du grand public de cette condition, au travers de films (e.g. Rain Man), de séries télé (e.g. The Big Bang Theory, Atypical), de romans (e.g. Le Bizarre Incident du chien pendant la nuit, ou Le théorème du homard), à l'instar du roman cité en début de chapitre. De manière générale, les personnages de ces fictions grand public correspondent aux critères diagnostiques de l'autisme, mais de façon archétypique (Nordahl-Hansen et al., 2018), ce qui peut à la fois sensibiliser le public à la question de l'autisme, tout en véhiculant une vision stéréotypée de cette condition. Dans une analyse de trois fictions récentes, Semino (2014) identifie en particulier qu'une interprétation littérale du langage figuré et des difficultés dans la gestion de l'informativité et de la pertinence constituent deux des trois grands types de difficultés pragmatiques représentées, qui sont utilisés par les auteurs pour attribuer aux protagonistes un 'style cognitif'<sup>12</sup> autistique.

Un rapport littéral au langage est en outre fréquemment rapporté par les personnes autistes elles-mêmes. Dans une série d'entretiens avec 18 adultes autistes, Müller et al. (2008) identifient les problèmes de communication comme un des défis rapportés par les interviewés, incluant des difficultés à saisir le sens implicite (mais aussi explicite) des énoncés, à faire des inférences sociales et émotionnelles, notamment lorsqu'il faut intégrer des indices non-verbaux et verbaux. Un de leurs participants rapporte : « *Je ne perçois pas le langage corporel ou le sens caché, ou ce que les gens disent vraiment quand ils disent ces trucs entre les lignes* »<sup>13</sup> (p. 179). Similairement, Josef Schovanec, témoigne ainsi de son autisme :

L'exemple le plus classique est peut-être la fameuse blague du contrôleur dans le train qui arrive et qui vous demande : « Est-ce que je peux voir votre billet ? » et vous vous répondez : « Non, vous ne pouvez pas le voir, il est dans ma poche ». Un exemple typique de non-compréhension sociale par la personne avec autisme, le plus souvent de la part des enfants dans ce cas précis. En effet, les adultes ont souvent appris à éviter un certain nombre d'erreurs élémentaires. Mais ils font d'autres gaffes du fait de la nature variable, fluctuante, inédite, imprévue des circonstances dans lesquelles nous nous mouvons. (Schovanec & Glorion, 2012/2013, p. 124)

De façon cohérente, on retrouve des questions ciblant ce type de difficultés dans le questionnaire de conversation construit par A. C. Wilson (2022) sur la base des

---

<sup>12</sup> Traduction de l'expression '*mind-style*' qui renvoie aux outils linguistiques qui peuvent être utilisés pour véhiculer une certaine représentation du monde, en l'occurrence autistique.

<sup>13</sup> Traduction personnelle de la citation originale : '*I can't pick up on body language or hidden meanings, or what people are really saying when they're saying that in-between-the-lines stuff.*'

expériences rapportées par les personnes autistes : ‘*Je suis confus quand les gens parlent indirectement*’, ‘*Je prends les choses littéralement*’, ‘*J’analyse ce que les gens veulent dire, parce que je ne comprends pas ou je crois que j’ai mal compris*’, etc. Montrant de bonnes qualités psychométriques, ce questionnaire pouvait en effet discriminer les participants autistes des participants non-autistes. Cette caractéristique du rapport au langage est également manifeste dans la description ironique que des personnes autistes peuvent faire des personnes non-autistes. Dans une démarche de normalisation de l’autisme, et afin d’attirer l’attention sur l’importance du point de vue dans lequel on se place pour qualifier un fonctionnement de normal ou non, la communauté autistique évoque en effet régulièrement les critères diagnostiques d’un *syndrome neurotypique*. Ce syndrome est défini comme un « *trouble neurobiologique caractérisé par un souci de préoccupations sociales, des délires de supériorité, et l’obsession de conformité* », touchant « *plus de 9 625 sur 10 000 personnes* ». Le détail des critères diagnostiques évoque entre autres une « *incapacité à ne pas interpréter les conversations avec autrui de manière illogique (en cherchant continuellement des doubles sens par exemple)* », une « *obsession marquée de comportements non-verbaux, vagues, indéfinissables* », décrivant encore qu’on n’observe « *pas de retard général du langage, mais [que] la récurrence du manque de précision et de clarté de la communication est significative sur le plan clinique* » (auteur inconnu, texte extrait d’un blog francophone<sup>14</sup>). Ainsi, cette description humoristique nous renvoie aux modèles sociaux du handicap et au problème de la double empathie de Milton (2012), mais renseigne également sur la perception de la communauté autistique de ses différences avec les personnes non-autistes, parmi lesquelles on retrouve l’utilisation d’une communication indirecte, implicite et vague.

L’étude expérimentale de ces différences est relativement plus récente, par rapport aux premières descriptions de l’autisme, et a connu un essor particulier à la rencontre du développement de la théorie de la pertinence, modélisant la communication comme basée sur la lecture des états mentaux d’autrui, et la théorie du défaut de théorie de l’esprit de l’autisme. Pour Happé (1993), les compétences communicatives des personnes autistes constituaient ainsi un « *test de la théorie de la pertinence* ». Dans une série de trois expériences, elle contraste la compréhension de comparaisons (e.g., ‘Il est comme un lion’), qui peuvent être comprises littéralement, à des métaphores (e.g. ‘C’est un lion’), pour lesquelles il faut se détacher (d’une partie) du sens littéral pour attribuer une intention de communication au locuteur, établissant un parallèle avec la compréhension de fausses croyances de premier ordre. Elle compare également ces phénomènes à la compréhension de l’ironie, qui impliquerait un niveau de complexité supérieur. En effet dans la théorie de la pertinence (voir [chapitre 1](#)), l’ironie repose sur la communication d’une *attitude* au sujet d’une *pensée* attribuée à autrui, et nécessiterait donc un degré de représentation supérieur ; aussi Happé (1993) lie-t-elle ce phénomène à la compréhension de fausses croyances de deuxième ordre.

---

<sup>14</sup> <https://spectredelautisme.blogspot.com/2014/11/syndrome-neurotypique.html>

Pour évaluer cette hypothèse, Happé (1993) teste des adolescents et des adultes autistes avec une batterie de théorie de l'esprit afin de constituer 3 groupes : 6 participants échouant aux épreuves de théorie de l'esprit de 1<sup>er</sup> et de 2<sup>ème</sup> ordre (groupe 'no-TOM'), 6 participants réussissant seulement les épreuves de théorie de l'esprit de 1<sup>er</sup> ordre (groupe 'TOM-1'), et 6 participants réussissant également les épreuves de théorie de l'esprit de 2<sup>ème</sup> ordre (groupe 'TOM-2'). Elle présente dans une première expérience à ces participants une tâche de complétion de phrases dans laquelle des énoncés doivent être complétés en sélectionnant parmi une série de propositions un mot constituant une comparaison ou une métaphore (ou pour une troisième condition contrôle, impliquant un synonyme). Le groupe de participants no-TOM réussit alors significativement moins bien les items impliquant des métaphores que les autres groupes, tandis que ce contraste n'était pas significatif pour les autres conditions. Les performances des groupes TOM-1 et TOM-2 ne différaient pas significativement. Dans une deuxième expérience, Happé compare la compréhension des métaphores et de l'ironie, en utilisant des histoires verbales dans lesquelles de tels énoncés sont insérés, et que les participants doivent expliquer. À nouveau, le groupe no-TOM réussit moins bien la condition métaphore que les deux autres groupes, tandis qu'aucune différence n'apparaît entre les groupes TOM-1 et TOM-2. Dans la condition ironie en revanche, le groupe TOM-1 obtient de moins bons résultats que le groupe TOM-2. Happé confirme ainsi expérimentalement que i) les participants autistes sont en difficulté avec le langage figuré, par rapport à leur compréhension d'expressions littérales équivalentes (comparaisons) ; et ii) que la nature de ces difficultés est associée à la mesure de leurs difficultés en théorie de l'esprit, apportant un soutien expérimental à la théorie de la pertinence.

Peu après, Surian et al. (1996) soutiennent cette idée en montrant qu'à la différence d'enfants avec un trouble du langage et d'enfants neurotypiques, des enfants autistes détectent peu des violations des maximes conversationnelles (informativité, vérité, pertinence, politesse), et que leurs performances sont associées à leur compréhension de fausses croyances. Ces articles ouvrent ainsi la voie à de nombreux travaux qui préciseront et nuanceront ces propositions. Nombre d'entre eux portent sur la compréhension des métaphores dans l'autisme et ses prédicteurs, comme nous le verrons dans la [partie 2](#). Plus récemment, cette investigation s'est élargie à une plus grande diversité de phénomènes pragmatiques, et en particulier à l'étude des implicatures scalaires, ce que nous détaillerons dans la [troisième partie](#) de ce chapitre.

## 2. Métaphores et autisme

### 2.1. Distinguer les particularités autistiques de difficultés verbales

L'étude princeps de Happé (1993) a suscité de nombreux travaux, notamment sur la base de critiques méthodologiques qui lui étaient faites. Outre la faible taille des échantillons inclus, les différents groupes de participants autistes évalués dans cette

étude n'avaient pas pu être appariés au niveau du fonctionnement intellectuel, verbal en particulier. Ainsi, les contrastes entre les performances en théorie de l'esprit recouvraient des contrastes entre les QI verbaux des participants, ce qui constituait une variable confondante limitant l'attribution des différences inter-groupes aux seules différences en théorie de l'esprit. Sur la base de cette critique, notamment, Norbury (2005) a proposé que dans l'autisme, la compréhension des métaphores soit expliquée par des difficultés verbales, plutôt qu'en théorie de l'esprit. Pour étayer cette proposition, elle a évalué avec une tâche de métaphores similaire des enfants de 8 à 15 ans avec des difficultés de communication, incluant des enfants avec des difficultés langagières formelles, des enfants avec des particularités autistiques, et des enfants présentant les deux caractéristiques. Un groupe contrôle d'enfants à développement typique était également inclus. Outre la compréhension des métaphores, la théorie de l'esprit et les compétences sémantiques des participants étaient aussi évaluées formellement. Les résultats montraient un effet principal de la condition (les items métaphores étaient moins réussis que les items contrôles) et un effet principal du groupe clinique (les enfants avec trouble du langage réussissant moins bien que les autres), mais pas d'interaction, donc pas de difficultés spécifiques à la condition métaphorique, relativement aux items contrôles. Le même pattern apparaissait en reclassifiant les enfants selon leurs performances en théorie de l'esprit comme dans l'étude de Happé (1993, no-TOM, TOM-1 et TOM-2). Les groupes no-TOM et TOM-1, en particulier, ne différaient pas spécifiquement sur les items métaphoriques, à la différence de l'étude de Happé (1993), mais avaient de moins bonnes performances que le groupe TOM-2 et le groupe contrôle. Dans les groupes cliniques, les performances en métaphores étaient significativement corrélées à la fois à la théorie de l'esprit et aux compétences sémantiques, mais pas à la symptomatologie autistique. La contribution respective de ces différents facteurs a donc été évaluée à l'aide de régressions hiérarchiques, ajustées sur la base d'un modèle prenant seulement en compte l'âge et les compétences lexicales de base (expliquant 22% de la variance totale). L'ajout au modèle des compétences sémantiques expliquait 12% de variance supplémentaire, tandis que l'ajout des performances en théorie de l'esprit n'améliorait pas les performances du modèle et que l'ajout de la mesure de la symptomatologie autistique ne l'améliorait que de 3%. Norbury (2005) a donc proposé que les représentations sémantiques soient le prédicteur le plus important pour la compréhension des métaphores, et que ce soit celui qui distingue des enfants avec trouble de la communication d'enfants à développement typique, plutôt qu'un statut autistique, ou que les performances en théorie de l'esprit qui seraient nécessaires mais non suffisantes.

Plusieurs limites méthodologiques empêchent toutefois de s'emparer complètement de cette contre-proposition, notamment la nature très verbale de la tâche de théorie de l'esprit utilisée (comme évoqué dans les chapitres précédents), et l'inclusion dans la mesure des compétences sémantiques d'un subtest évaluant la compréhension d'idiome, induisant ainsi une variable confondante importante. Cependant, cette opposition avec les propositions de Happé a abouti à de nombreuses autres études,

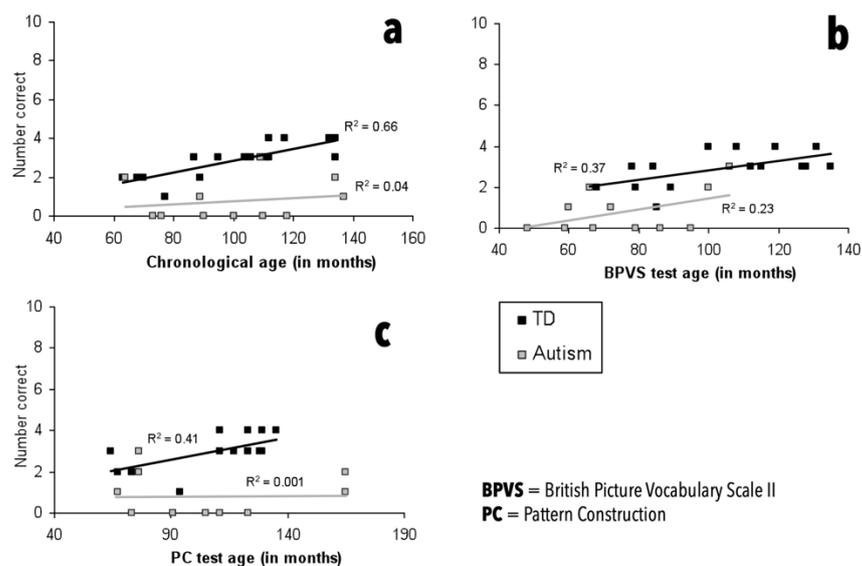
visant premièrement à déterminer les prédicteurs des difficultés que peuvent rencontrer ces enfants, puis les prédicteurs du développement typique des métaphores (comme décrit au [chapitre 2](#)). D'autre part, ces données questionnaient l'attribution même de difficultés en compréhension de métaphores aux enfants autistes, ce que de nombreuses études ont par la suite exploré.

Rundblad et Annaz (2010) ont notamment utilisé une méthodologie particulièrement sensible au développement, en évaluant (notamment) la compréhension des métaphores de 11 enfants autistes de 5 à 11 ans et de 17 enfants à développement typique, afin de comparer leurs trajectoires développementales respectives. Elles ont pour ce faire utilisé une tâche dans laquelle de courtes histoires étaient lues aux enfants, accompagnées de quelques images illustrant le récit, et se terminant par une métaphore conventionnelle. Suite à chaque histoire, une question ouverte sur ce qu'un personnage avait vu était posée à l'enfant, nécessitant d'identifier le référent de la métaphore (par exemple, « Qu'a vu Stuart ? » après qu'une foule de gens ait été décrite comme une « marée »<sup>15</sup>). Des questions plus précises pouvaient être posées si l'enfant ne répondait pas. Les auteurs caractérisaient également l'âge mental non-verbal des participants avec un test de reproduction de figures, et leur âge mental verbal à l'aide d'un test de vocabulaire. Leurs résultats montrent que, si les performances des enfants à développement typique progressent avec l'âge, c'est beaucoup moins le cas des enfants autistes dont les performances sont nettement inférieures, avec un effet plancher pour ce groupe (figure 3-4, a). Par ailleurs, les trajectoires développementales des deux groupes restaient nettement dissociées lorsque leurs performances étaient régressées non en fonction de leur âge chronologique mais de leur âge mental verbal (figure 3-4, b) ou non-verbal (figure 3-4, c). Ces observations suggèrent ainsi que les nettes difficultés des participants autistes pourraient bien être attribuées à leur groupe diagnostic, et non à des différences au niveau du raisonnement ou du vocabulaire.

Van Herwegen et Rundblad (2018) obtiennent des résultats similaires en adaptant cette tâche avec des métaphores nouvelles plutôt que conventionnelles : alors que les participants neurotypiques montraient des meilleures performances lorsqu'ils étaient plus âgés, ce n'était pas le cas des participants autistes, dont les performances ne semblaient pas liées à l'âge. Dans une seconde expérience, Van Herwegen et Rundblad (2018) réévaluent 8 participants autistes parmi les plus jeunes de l'échantillon initial, à distance de cette première évaluation (après un délai moyen de 34 mois). Ces données longitudinales montrent toutefois que la compréhension des métaphores progresse avec l'âge chez ces enfants, témoignant d'un développement atypique ou différé plutôt qu'absent, malgré le fait que l'âge chronologique soit un mauvais prédicteur des performances dans les méthodologies transversales.

---

<sup>15</sup> Traduction personnelle de la question « What did Stuart see ? » après une métaphore décrivant une foule de gens avec le mot 'flood' [inondation].



**Figure 3-4 :** Performances en compréhension de métaphores des enfants autistes et neurotypiques (TD) en fonction de l'âge chronologique (a), de l'âge mental verbal (b) et de l'âge mental non-verbal (c) dans l'étude de Rundblad et Annaz (2010 ; figures extraites de la publication originale)

Cette différence n'est pas toujours aussi évidente mais est observée dans de nombreuses études, comme le révèle notamment la méta-analyse de Kalandadze et al. (2018) pour le langage figuré en général (montrant même une différence plus nette sur la compréhension des métaphores qu'en compréhension de l'ironie). Plus spécifiquement pour la compréhension des métaphores, c'est également ce que montrent deux autres méta-analyses de Kalandadze et al. (2019) ou de Morsanyi et al. (2020), incluant à la fois des études chez les enfants et les adultes. Ces méta-analyses mettent en évidence des effets modérés à larges en faveur des participants contrôles non-autistes. Cette dernière revue synthétise aussi les résultats de 5 études rapportant les temps de réponse de participants dans des tâches de métaphores et met également en évidence un effet de taille similaire en faveur des participants neurotypiques. Parmi les quelques études employant ce type de mesures, Morsanyi et al. (2022) ont récemment proposé à des adultes autistes ou non une tâche de décision métaphorique, dans laquelle les participants devaient déterminer si une expression présentée à l'écrit était métaphorique ou non. Alors qu'aucune différence entre les groupes n'apparaissait dans les réponses en elles-mêmes, les participants autistes répondaient plus lentement que les participants contrôles, suggérant un traitement atypique engagé par la tâche.

Ces trois méta-analyses relèvent toutefois l'importante hétérogénéité de la littérature et de la taille des effets rapportés. Les trois revues investiguent des facteurs en rapport avec l'hypothèse de Norbury (2005) selon laquelle les difficultés des personnes autistes seraient expliquées par des fragilités verbales. Kalandadze et al. (2018) montrent que la différence inter-groupe est robustement observée lorsque les participants sont appariés en termes de fonctionnement intellectuel, mais pas dans les études qui appariant les groupes en termes de compétences langagières, ce qui va en faveur de l'hypothèse de Norbury. Toutefois, la méta-analyse met également en

évidence une différence entre les groupes lorsqu'ils sont appariés sur ces deux facteurs conjointement, ce qui empêche de conclure sur cette question. En considérant seulement des études incluant des groupes appariés en âge et en QI (Quotient Intellectuel) verbal ou QI total, Morsanyi et al. (2020) confirment cette différence entre les groupes, mais observent que cette différence tend à se réduire lorsque le QI verbal des participants est élevé (avec des effets larges en cas de QI dans la norme à des effets plus modérés en cas de QI verbal supérieur à la norme). Kalandadze et al. (2019) explorent en outre la possibilité que la variabilité des effets dépende des propriétés des tâches utilisées, mais la littérature est trop hétérogène et réduite pour pouvoir explorer cette question statistiquement. Les auteurs remarquent toutefois qualitativement que les effets les plus prononcés sont observés dans des études utilisant des tâches d'explications verbales libre. Or, comme décrit au [chapitre 2](#), ces tâches impliquent également des compétences métalinguistiques et langagières expressives qui pourraient mettre en difficultés les participants autistes au-delà de leur compréhension des métaphores seulement. Similairement, la conventionnalité des métaphores utilisées est susceptible de modifier les différences constatées, comme dans le fonctionnement typique (voir [chapitre 1](#)) : une métaphore conventionnelle est plus susceptible d'être traitée comme un item de la langue en tant que tel, donc plus associée aux connaissances langagières, tandis que la compréhension d'une métaphore nouvelle devrait plus probablement être entravée par des difficultés en théorie de l'esprit. Toutefois peu d'études ont investigué l'effet de ce paramètre, et leurs résultats divergent (Kalandadze et al., 2019). De même, certaines études ont utilisé des métaphores décontextualisées, et d'autres insérées dans un contexte visuel ou verbal plus ou moins riche, mais les effets de telles manipulations dans cette population sont encore largement inconnus (Kalandadze et al., 2019).

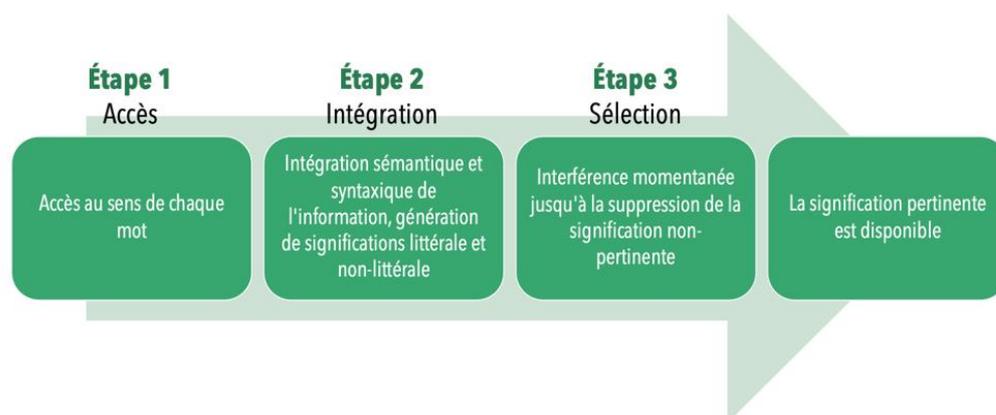
En outre, les compétences des participants autistes avec les métaphores ne doivent pas être considérées de manière figée, d'une part parce que malgré ces différences inter-groupes les performances des participants autistes augmentent bien avec l'âge, comme montré par exemple dans l'étude longitudinale de Van Herwegen et Rundblad (2018) citée plus haut. D'autre part, quelques études semblent indiquer que ces compétences peuvent spécifiquement être entraînées chez des enfants autistes (Melogno & Pinto, 2022). Ces entraînements s'appuient sur différents ancres théoriques allant d'approches comportementalistes (e.g. G. T. Lee et al., 2019) ou plus psycholinguistiques (e.g. Mashal & Kasirer, 2011), et leurs résultats positifs, bien qu'encore préliminaires, indiquent que les inférences pragmatiques permettant la compréhension des métaphores ne doivent pas être considérées comme *intrinsèquement* déficientes dans l'autisme.

Si à première vue les résultats expérimentaux semblent donc confirmer le déficit en compréhension des métaphores associé à l'autisme, un examen attentif des données pousse donc à nuancer ce constat et empêche encore de tirer des conclusions claires, en contraste avec les représentations très franches associées à l'autisme ([section 1.1](#)). En effet, les résultats des études apparaissent modulés par un grand nombre de

paramètres, liés aux caractéristiques des participants eux-mêmes ou des dispositifs expérimentaux très variés qui sont utilisés. De plus, ces difficultés potentielles ne sont pas figées, et pourraient être attribuées non pas à la condition autistique des participants, mais à des fragilités verbales qui sont fréquemment associées à cette condition, remettant en question la spécificité de cette caractéristique (Gernsbacher & Pripas-Kapit, 2012). Une approche plus nuancée serait de considérer que l'autisme serait associé à un traitement *atypique* des métaphores, plutôt qu'une simple *difficulté* à les comprendre. L'observation d'effets sur les temps de réponses plutôt que sur les réponses elles-mêmes à des stimuli impliquant des métaphores, ou d'activités atypiques au niveau cérébral par rapport à des participants non-autistes (e.g. Gold et al., 2010) est particulièrement en faveur de cette position.

## 2.2. Les mécanismes de compréhension des métaphores

Quelques études ont par ailleurs spécifiquement évalué les mécanismes sous-jacents à la compréhension des métaphores dans l'autisme. Chouinard et Cummine (2016) ont par exemple recherché un effet d'interférence métaphorique dans cette population, renseignant sur différentes étapes du modèle d'interprétation des métaphores développé par Glucksberg et al. (1982) (cf. [chapitre 1](#) et figure 3-5).



**Figure 3-5 :** Représentation des étapes cognitives de la compréhension d'une métaphore d'après le modèle de Glucksberg et al. (1982), dont l'interférence métaphorique témoignerait (figure traduite de Chouinard et al., 2017)

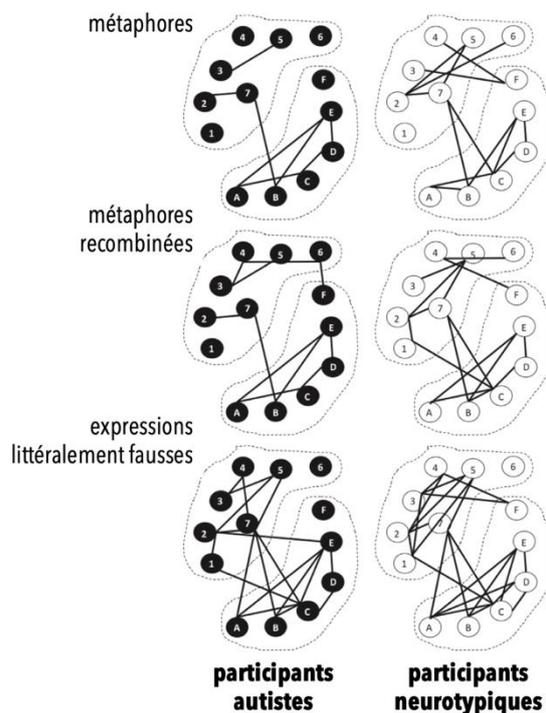
Chouinard et Cummine (2016) ont ainsi évalué 13 participants autistes sans handicap intellectuel ni trouble du langage et 12 participants contrôles neurotypiques, appariés en âge, QI non-verbal et compétences sémantiques, et leur ont proposé une tâche de décision de phrase, dans laquelle des participants devaient juger si des expressions proposées étaient *littéralement* vraies ou non. Comme l'étude originale (mais à l'oral et non à l'écrit), les stimuli incluent des phrases littéralement vraies, littéralement fausses, des métaphores, et des métaphores recombinaées, dans lesquelles les termes de différentes métaphores sont combinées pour constituer d'autres expressions littéralement fausses. Alors que l'appariement soigné des groupes doit assurer leur équivalence au niveau de la première étape (accès au sens de chaque mot), la présence

d'un effet d'interférence métaphorique doit témoigner de la génération automatique d'une signification non-littérale (étape 2) tandis que la taille de cet effet et les réponses à la condition critique doivent renseigner sur l'étape de suppression de la signification non-pertinente.

Chouinard et Cummine (2016) n'observent pas de différence entre les groupes au niveau des réponses données, dans chaque condition. Au sein de chaque groupe toutefois, les participants contrôles répondaient autant 'Faux' dans le cas des métaphores que dans les conditions contrôles où cette réponse était attendue, tandis que les participants autistes tendaient à commettre plus d'erreurs spécifiquement sur les métaphores, suggérant des difficultés dans la gestion de l'interférence métaphorique. Au niveau des temps de réponse, un effet d'interférence métaphorique était bien observé dans les deux groupes, marqué par des réponses plus lentes dans la condition métaphorique que pour les expressions contrôles fausses (en cohérence avec les observations de Hermann et al. [2013] avec des stimuli écrits). Aucune différence n'était observée dans la taille de cet effet. Ces résultats confirment donc que l'étape de génération automatique d'une signification métaphorique serait bien présente dans l'autisme, comme chez des participants neurotypiques, et que les différences associées à cette condition pourraient plutôt intervenir dans l'étape ultérieure de sélection du candidat le plus pertinent. Les résultats renseignant sur cette étape sont plus contrastés, avec des observations qui soutiennent cette hypothèse, puisque seuls les participants autistes réalisent plus d'erreurs dans la condition métaphorique que les autres, ce qui indiquerait une interférence plus forte de la signification métaphorique de l'expression. Les temps de réponse, en revanche, ne soutiennent pas cette hypothèse puisqu'aucune différence dans les tailles d'effet d'interférence n'a pu être observée. Les effectifs des groupes, toutefois, rendaient cette étude peu puissante pour détecter de tels effets.

Par ailleurs, cette étude était réalisée en situation d'imagerie cérébrale fonctionnelle, dont les résultats sont rapportés par Chouinard et al. (2017). En contrastant les conditions métaphores et métaphores recombinaisons (condition contrôle) pour isoler les processus de sélection de la signification pertinente, ces données révèlent l'implication de zones cérébrales différentes entre les deux groupes. Alors que chez les participants neurotypiques la condition métaphore est associée à une suractivation dans le cortex cingulaire antérieur gauche et le cortex cingulaire postérieur droit, un réseau plus étendu était observé chez les participants autistes, avec une suractivation des gyri frontaux supérieur gauche et inférieur droit, du thalamus et du lobule pariétal supérieur. Par ailleurs, une analyse de la connectivité fonctionnelle lors des différentes conditions (figure 3-6) mettait en évidence une moindre connectivité chez les participants autistes que les participants neurotypiques dans la condition métaphore, impliquant la sélection de la signification pertinente, en particulier une moindre connectivité cortico-corticale. Moins de différences entre les groupes étaient observées dans les conditions contrôles (figure 3-6). Ces observations renforcent donc l'hypothèse d'un traitement cérébral et cognitif atypique des métaphores, et plus particulièrement dans l'étape de sélection de la signification pertinente d'une

expression et de la suppression de la signification non-pertinente, plutôt qu'à l'étape de génération des différentes significations.



**Figure 3-6 :** Représentation visuelle des analyses de connectivité extraite de Chouinard et al. (2017), représentant les liens significatif (traits pleins) entre les différentes régions d'intérêt considérées, incluant des régions corticales (1 à 7) et sous-corticales (A à F). 1 = gyrus temporal médian, 2 = gyrus temporal supérieur, 3 = cortex préfrontal dorsolatéral, 4 = cortex cingulaire antérieur, 5 = gyrus frontal inférieur, 6 = cortex pariétal postérieur droit, 7 = insula ; A = noyau subthalamique, B = putamen, C = thalamus, D = globus pallidus interne, E = globus pallidus externe, F = noyau caudé

Ces résultats indiquent donc que l'autisme pourrait être associé à des différences dans le traitement des métaphores, indépendamment du niveau langagier des participants. Ces différences font écho à d'autres hypothèses explicatives des difficultés classiquement associées à l'autisme, outre les deux candidats principaux que sont la théorie de l'esprit et le langage. En effet, les inférences pragmatiques des personnes autistes ont été mises en rapport avec d'autres particularités de leur fonctionnement cognitif au niveau exécutif (Deliens et al., 2018; Kissine, 2012) et de la cohérence centrale (Happé & Frith, 2006). Kissine (2012) propose par exemple que la moindre flexibilité cognitive associée à l'autisme (e.g. Prior & Hoffmann, 1990; mais voir aussi Leung & Zakzanis, 2014) puisse expliquer certaines difficultés pragmatiques des personnes concernées. En effet, comprendre ces phénomènes nécessite d'alterner de façon fluide entre sa perspective et celle du locuteur (tout comme répondre à une épreuve de fausses croyances). Les inférences pragmatiques reposent par ailleurs souvent sur la prise en compte d'informations multiples et contextuelles (Frith, 1989). Or, on retrouve dans l'autisme une tendance générale à favoriser le traitement local, plutôt que de l'aspect global de l'information. Ce style cognitif atypique des personnes autistes (Frith & Happé, 1994) pourrait ainsi également contribuer à expliquer leurs

particularités dans les inférences pragmatiques (Happé & Frith, 2006). Sans qu'il ne soit nécessaire de détailler plus avant le contenu et les motivations de ces hypothèses pour ce travail de thèse, celles-ci illustrent bien la diversité des explications possibles de particularités autistiques, qui ne sont pas (toujours) exclusives. Vulchanova et al. (2015) proposent d'ailleurs que l'intégration de différentes informations et de processus cognitifs différents puisse en elle-même constituer un point de divergence entre l'autisme et le fonctionnement neurotypique.

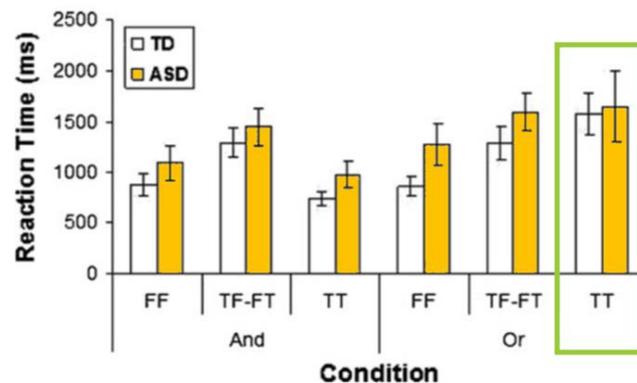
Malgré une certaine convergence de surface de nombreux travaux, la littérature sur les métaphores révèle donc une absence de consensus sur la nature des particularités autistiques vis-à-vis des inférences pragmatiques, ce qui fait écho à la multiplicité des explications de l'autisme de manière générale. Quoi qu'il en soit, ces données mettent par ailleurs en doute toute proposition selon laquelle la compréhension pragmatique est altérée de façon générale dans l'autisme. Ce tableau est d'autant plus nuancé par l'investigation plus récente des implicatures scalaires chez des participants autistes.

### 3. Implicatures scalaires et autisme

#### 3.1. Un domaine pragmatique préservé ?

Pour les raisons évoquées dans les deux premières parties de ce chapitre, les premières études évaluant la compréhension d'implicatures scalaires par des participants autistes, en 2009 et 2010, s'attendaient à observer des performances moindres par rapport à des contrôles neurotypiques. Les résultats de Pijnacker et al. (2009) puis de Chevallier et al. (2010) sont pourtant allés à l'encontre de cette hypothèse. Chacune de ces études a employé un paradigme classique de vérification de phrases, avec le quantifieur '*certain*' et la disjonction '*ou*', de façon complémentaire. Pijnacker et al. (2009) ont inséré ces termes scalaires dans des phrases écrites, pouvant être sous-informatives compte-tenu des connaissances encyclopédiques des participants (e.g. '*certain* moineaux sont des oiseaux' ou '*les zèbres ont des rayures noires ou blanches*'). Ils les ont présentés à des adultes autistes sans handicap intellectuel ni trouble du langage, appariés à des adultes neurotypiques en âge, QI verbal et QI non-verbal. Chevallier et al. (2010) se sont quant à eux focalisés sur la disjonction '*ou*' et l'ont insérée dans des phrases orales, pouvant être sous-informatives compte-tenu d'un contexte visuel qui leur était associé (par exemple '*il y a un cheval ou un chien*', avec une image représentant un cheval *et* un chien). Ils ont évalué le jugement de ces phrases par des adolescents de 10 à 16 ans, autistes ou non, appariés en âge et en niveau de vocabulaire. Cependant dans ces deux études, aucune différence n'a été observée entre les groupes autistes et neurotypiques, ni dans l'acceptation ou le rejet des phrases sous-informatives, ni dans les phrases contrôles. Par ailleurs, chaque étude mesurait également les temps de réponse (ou temps de lecture), qui révélaient les mêmes patterns de temps de réponse sur les énoncés critiques ou contrôles pour les deux groupes (voir figure 3-7). Ces résultats suggèrent donc que les similarités observées au niveau

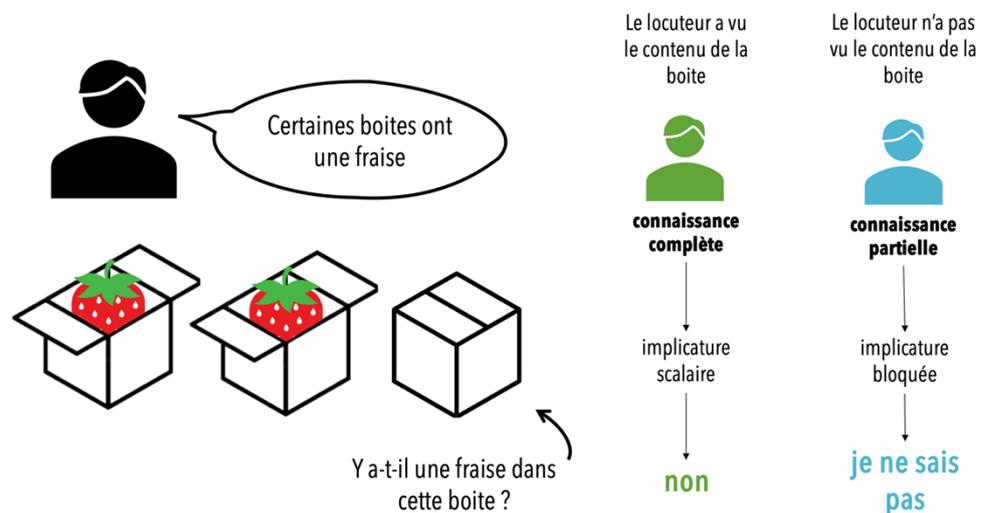
comportemental ne sont pas seulement des similarités de surface, mais qu'elles reflètent des similarités dans le traitement même de ces énoncés.



**Figure 3-7 :** Temps de réponse observé par Chevallier et al. (2010) chez les participants autistes (ASD, en jaune) ou non (TD, en blanc) dans la condition critique (encadrée en vert) ou les conditions contrôles (couleurs et encadré ajoutés à la figure originale)

Ce résultat dans des tâches de vérification de phrases a de nouveau été répliqué chez des enfants (Andrés-Roqueta & Katsos, 2020; Schaeken et al., 2018; Su & Su, 2015) et des adolescents (Hochstein et al., 2017), et peut donc être considéré comme robuste, suggérant que les implicatures scalaires pourraient constituer un domaine pragmatique qui serait typique dans l'autisme. Toutefois, d'autres résultats nuanciant cette conclusion étaient parfois associés à ces réplifications. Andrés-Roqueta et Katsos (2020) ont notamment mis en lumière l'importance de l'appariement des groupes, puisque les enfants autistes de cette étude réalisaient des performances équivalentes à celles d'enfants à développement typique appariés en niveau de langage, mais des performances plus pauvres que des enfants appariés en âge chronologique (rappelant les discussions des performances en compréhension de métaphores évoquées ci-dessus). Cependant Schaeken et al. (2018) n'observent pas cette différence, et les enfants autistes qu'ils évaluent tendent même à rejeter les énoncés sous-informatifs *plus fréquemment* que des enfants à développement typique appariés en âge. Cette étude introduisait également dans une deuxième expérience une tâche de jugement ternaire, évoquée au [chapitre 2](#), faisant apparaître une différence entre ces groupes. Tandis que les enfants neurotypiques préféraient la réponse intermédiaire ('je suis un peu d'accord') pour évaluer des phrases sous-informatives (cf. [chapitre 2](#)), les enfants autistes conservaient plus souvent des rejets complets ('pas du tout d'accord') de ces expressions. Les auteurs évoquent ainsi des jugements *dichotomisés* associés à l'autisme, qui pourraient révéler des particularités dans la gestion de l'informativité, habituellement masquées dans les tâches de jugement binaires classiques. Cette différence disparaissait toutefois lorsque les performances des participants autistes étaient comparées à des participants contrôle appariés non plus en âge, mais en niveau de QI (ce groupe incluant également des enfants avec d'autres troubles du neurodéveloppement), ce qui met en doute qu'une telle particularité soit spécifique à l'autisme.

Après avoir répliqué le résultat classique avec une tâche de jugement binaire, Hochstein et al. (2017) introduisent également une modification dans le paradigme qui révèle des particularités chez les adolescents autistes évalués. Dans une tâche dite de ‘connaissance partielle’ inspirée de Bergen et Grodner (2012), les auteurs introduisent un personnage produisant des énoncés avec le quantifieur ‘certaines’ (ou d’autres quantifieurs contrôle) au sujet du contenu de trois boîtes. Le contenu de deux boîtes est ensuite dévoilé au participant, et il lui est demandé de se prononcer sur le contenu de la troisième boîte, étant donné le contenu des deux autres et l’énoncé du personnage (voir figure 3-8). Par exemple, le personnage inspecte le contenu des trois boîtes, puis déclare que ‘*certaines boîtes ont une fraise*’. Si les deux premières boîtes contiennent une fraise, un participant pourrait faire une interprétation pragmatique de l’énoncé et inférer que toutes les boîtes ne contiennent pas de fraise, et donc que la troisième boîte n’en contient pas. À la différence d’une tâche de vérification de phrase, cette tâche s’approche donc plus de conditions naturelles dans laquelle un état du monde doit être inféré à partir d’un énoncé. Cette condition dans laquelle le personnage avait regardé dans l’ensemble des boîtes (‘connaissance complète’) est comparée à une condition dans laquelle le personnage ne regarde que le contenu des deux premières boîtes, avant de déclarer que ‘*certaines boîtes ont une fraise*’. Dans ce contexte (‘connaissance partielle’), la dérivation de l’implicature scalaire n’est pas attendue, car le locuteur n’a pas la connaissance du contenu de toutes les boîtes. Lorsqu’il est demandé au participant si la troisième boîte contient une fraise, la réponse attendue est donc ‘*je ne sais pas*’ plutôt que ‘*non*’ (voir figure 3-8). Par ailleurs, la tâche inclut des conditions contrôles notamment avec le quantifieur ‘toutes’.



**Figure 3-8 :** Illustration du paradigme de connaissance partielle d'Hochstein et al. (2017) avec les réponses attendues aux items critiques (en couleur) en fonction de la condition

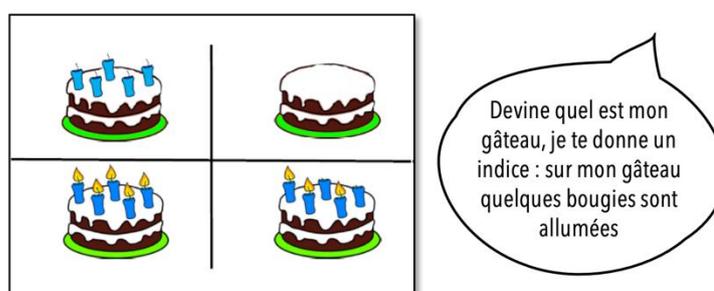
Les réponses étaient très bonnes de façon générale dans les items contrôles. Pour ce qui est des items critiques, dans la condition ‘connaissance complète’ les adolescents autistes produisaient le plus souvent la réponse attendue (‘non’, environ 70% des réponses), dans la même mesure que le groupe contrôle neurotypique. Comme suggéré

par les tâches de jugement binaire, ils montrent donc également dans cette situation plus naturelle une interprétation pragmatique et typique de termes scalaires. Dans la condition ‘connaissance partielle’ en revanche, les participants neurotypiques donnent la réponse attendue (*‘je ne sais pas’*) dans la plupart des cas (environ 90%) tandis que c’est beaucoup moins le cas des participants autistes (environ 40% des réponses), qui répondent le plus souvent *‘non’* (47% des réponses environ). Ces participants tendent donc à interpréter ‘certaines’ comme ‘certaines-mais-pas-toutes’ nettement plus fréquemment que les participants contrôles, dans cette condition où les connaissances du locuteur dissuadent les participants neurotypiques d’une telle inférence. En d’autres termes, les participants neurotypiques modulent spontanément leur interprétation des quantifieurs en fonction des connaissances qu’ils attribuent au locuteur, tandis que les adolescents autistes beaucoup moins, témoignant plus souvent d’une interprétation dite pragmatique, indépendamment du contexte épistémique. Ce résultat étend donc les conclusions des travaux précédents à des situations plus naturelles, mais les module aussi fortement en montrant une moindre prise en compte spontanée des informations épistémiques dans ces processus. Sur ce dernier point toutefois, cette étude présente des limitations importantes. Premièrement, les auteurs pointent eux-mêmes que les adolescents autistes étaient comparés à des adultes neurotypiques, donc plus âgés, et non appariés en QI ou en niveau de langage. Il est donc possible que cette différence explique les divergences comportementales des groupes dans la condition ‘connaissance partielle’. De plus, la réponse attendue dans les items critiques de cette condition était *‘je ne sais pas’*, ce qui n’était le cas d’aucune des conditions contrôles. Or, il est possible que cette réponse en particulier ait pu mettre les adolescents concernés en difficulté par rapport aux adultes neurotypiques. Dans l’autisme (ou chez des participants moins âgés que dans le groupe contrôle), une simple difficulté à utiliser cette réponse (qui renvoie elle-même à un état épistémique, plus que oui ou non) pourrait ainsi expliquer le pattern observé. Le fait qu’un expérimentateur pose une question sur le contenu d’une boîte, dans le cadre d’une recherche, peut par exemple sous-entendre qu’une réponse à cette question existe, et décourager l’expression d’une ignorance, par rapport à d’autres contextes, ce à quoi des adolescents autistes pourraient être différemment sensibles, par rapport à des adultes neurotypiques.

Enfin, deux études plus récentes ont introduit d’autres modalités d’évaluation des implicatures scalaires à l’aide de questions à choix multiples, qui tendent à montrer des différences entre des participants autistes et des participants neurotypiques, à la différence de tâches de jugement binaire. Pastor-Cerezuela et al. (2018) comparent ainsi un groupe d’enfants et préadolescents autistes (6 à 13 ans) à des participants neurotypiques appariés en âge ou en niveau de vocabulaire avec un test d’implicatures conversationnelles variées, en Espagnol. Dans ce test, des implicatures sont présentées à l’écrit aux participants, qui doivent sélectionner parmi 3 propositions la paraphrase qui correspond le mieux à l’énoncé cible. Un (unique) exemple d’implicature scalaire est intégré au sein d’un groupe de 5 implicatures de type Q (selon la typologie de Levinson [2000]) : les participants doivent choisir la paraphrase la plus appropriée à

l'énoncé '*Certains invités sont venus à la fête de Maria*, parmi les propositions suivantes : (a) Tous les gens que Maria avait invités sont venus (interprétation logique), (b) Tous les gens que Maria avait invités ne sont pas venus (interprétation pragmatique attendue), et (c) Exactement trois invités sont venus. Sur l'ensemble de ces implicatures, les participants autistes réalisent des performances plus faibles que les deux groupes contrôles, et répondent plus lentement (la nature des erreurs des participants n'est pas rapportée). Ce résultat suggère que cette modalité d'évaluation pourrait révéler des fragilités autistiques avec les implicatures scalaires, mais il n'est pas possible d'isoler les performances sur ce phénomène seulement, et ce test n'incluait pas d'items contrôles. Il n'est donc pas possible d'exclure que les participants autistes aient été en difficulté avec la tâche en général (par exemple simplement en lecture), plutôt qu'avec les phénomènes ciblés par les auteurs.

Mazzaggio et al. (2021) ont également proposé une épreuve de choix multiples mais à partir d'images, à des enfants autistes de 4 à 9 ans et des enfants neurotypiques appariés en âge, QI et vocabulaire (mais divergeant en compréhension syntaxique). Dans cette tâche, il est demandé aux enfants de pointer une image parmi quatre, selon une consigne donnée (voir figure 3-9), cette consigne pouvant inclure un énoncé suscitant une implicature scalaire (par exemple '*Sur mon gâteau, quelques bougies sont allumées*'). Les propositions de réponse incluent des représentations d'une interprétation pragmatique de l'énoncé, d'une interprétation logique, et deux distracteurs. Cette tâche incluait des implicatures ad-hoc, également basées sur l'informativité (mais sur une échelle construite à partir d'un contexte plutôt que dans la langue), et des conditions contrôles pour ces deux phénomènes pragmatiques, incluant le quantifieur 'toutes' pour les implicatures scalaires.



**Figure 3-9 :** Exemple d'item critique avec une implicature scalaire, extrait de Mazzaggio et al. (2021) : la réponse attendue est l'image en bas à droite, tandis que l'image en bas à gauche est une réponse possible en cas d'interprétation logique (support visuel extrait de la publication originale, consigne orale traduite depuis l'italien)

Avec cette tâche, des analyses exploratoires non-paramétriques tendaient à mettre en évidence des difficultés des participants autistes sur les implicatures scalaires, par rapport aux enfants à développement typique. Toutefois, l'analyse principale révélait seulement que les enfants autistes réalisaient de moindres performances de façon générale (sur tous les types d'items), mais ne montraient pas de difficultés spécifiques

sur les items impliquant les implicatures scalaires par rapport aux items contrôles (aucune interaction entre le groupe et la condition n'apparaît).

### **3.2. Prédicteurs des implicatures scalaires dans l'autisme**

Parmi ces études, quelques-unes d'entre-elles ont également évalué d'autres compétences cognitives qui pourraient soutenir la compréhension des implicatures scalaires des participants autistes, telles que l'intelligence verbale ou non-verbale, les compétences langagières formelles ou la théorie de l'esprit. Ces résultats apparaissent toutefois très contradictoires. Certaines études observent des associations entre les performances dans des tâches d'implicatures scalaires et le QI verbal (Chevallier et al., 2010; Pijnacker et al., 2009) ou les compétences langagières formelles (Andrés-Roqueta & Katsos, 2020), tandis que d'autres études n'ont pas observé d'association avec le langage (Hochstein et al., 2017; Mazzaggio et al., 2021; Pijnacker et al., 2009), mettant plutôt en évidence des relations avec la théorie de l'esprit (Mazzaggio et al., 2021) ou non (Hochstein et al., 2017). Il apparaît donc difficile de tirer des conclusions face à une telle hétérogénéité, d'autant plus que très peu d'études ont évalué ces facteurs conjointement et de façon comparable.

Par ailleurs, une fraction de ces études a recherché de telles associations également au sein des groupes neurotypiques contrôles évalués, afin de comparer la nature des prédicteurs selon les groupes. Chevallier et al. (2010) ont ainsi observé une relation entre le QI verbal et les implicatures scalaires chez les adolescents autistes mais pas chez les contrôles neurotypiques. À l'inverse, Mazzaggio et al. (2021) rapportent des performances prédites par la théorie de l'esprit au sein du groupe d'enfants autistes mais par les compétences langagières formelles chez les enfants à développement typique. Si ces résultats suggèrent tout deux l'existence de particularités autistiques, non dans les performances des participants mais dans les mécanismes sous-jacents qui les déterminent, leurs directions contradictoires incitent fortement à la prudence dans leur interprétation.

À l'issue de cette revue de littérature, à notre connaissance exhaustive, il apparaît donc qu'aucune différence entre des participants autistes et neurotypiques, enfants ou adultes, n'apparaît dans des tâches de vérifications de phrases avec des implicatures scalaires, ni dans les réponses des participants, ni dans leurs temps de réponse. Ce résultat robuste pourrait indiquer que les implicatures scalaires constituent un domaine pragmatique préservé dans l'autisme. Quelques résultats suggèrent toutefois que cette modalité d'évaluation pourrait dissimuler des particularités autistiques dans le traitement des implicatures scalaires, mais leurs limites méthodologiques ne permettent pas pour l'heure d'adopter cette position. Bien qu'elles appellent d'autres investigations, l'ensemble de ces observations interroge donc les modèles actuels de l'autisme, intrinsèquement associés à des difficultés pragmatiques générales. Par ailleurs, la recherche sur d'autres conditions neurodéveloppementales a également fait apparaître des difficultés pragmatiques en dehors de l'autisme, notamment dans le

cadre du trouble développemental du langage, dans lequel les difficultés pragmatiques sont considérées comme alignées sur les difficultés verbales (Bishop et al., 2017; Norbury, 2005; Norbury et al., 2014). Ces observations nous poussent également à réexaminer les modèles pragmatiques en général.

## 4. Compétence unique ou processus multiples

Le contraste apparent entre des comportements pragmatiques atypiques (la compréhension du langage figuré et notamment des métaphores) et d'autres typiques dans l'autisme (implicatures scalaires) contribue à fragmenter le bloc monolithique d'une 'compétence pragmatique' (e.g. Geurts et al., 2019) et alimente une discussion sur la nature même de ces processus. Andrés-Roqueta et Katsos (2017) en particulier s'appuient sur l'idée que la relation entre pragmatique, langage et théorie de l'esprit ne soit pas fixe, mais modulée selon les situations (en cohérence avec plusieurs propositions théoriques, e.g., Kissine, 2012; Recanati, 2004; Sperber, 1994), pour expliquer ces contrastes. Ils introduisent une dichotomie opposant pragmatique linguistique et pragmatique sociale<sup>16</sup>. La *pragmatique linguistique* désigne alors les phénomènes qui peuvent être compris sur la base des compétences langagières formelles et de 'normes pragmatiques', tandis que la *pragmatique sociale* mobiliserait également les habiletés à se représenter les intentions, désirs et croyances d'autrui, donc la théorie de l'esprit. Cette distinction ne constitue pas une catégorisation absolue des phénomènes pragmatiques en tant que tels, mais plutôt de situations, selon les compétences qu'elles mobilisent. Ainsi, des implicatures scalaires sont typiquement considérées comme un exemple de pragmatique linguistique, mais pourraient dans certains contextes nécessiter le recrutement de la théorie de l'esprit. À l'inverse, le langage figuré, et en particulier l'ironie, est typiquement considéré comme un exemple de pragmatique sociale. Andrés-Roqueta et Katsos proposent que l'autisme soit caractérisé par une altération de la pragmatique sociale, en lien avec les particularités en théorie de l'esprit décrite à la [section 1.1](#), mais pas de la pragmatique linguistique, chez les individus sans trouble du langage.

Il faut noter que ce contraste émerge de la comparaison d'études multiples, chacune se focalisant sur certains phénomènes, mais que très peu d'entre-elles, à notre connaissance, l'ont recherché au sein d'un même échantillon. C'est toutefois le cas de Deliens et al. (2018) qui observent chez des adultes autistes une moindre compréhension de l'ironie que des adultes neurotypiques, mais des performances similaires en compréhension de requêtes indirectes. Certains travaux confirment par la suite que les requêtes indirectes pourraient être comprises par des enfants autistes, mais mettent également en évidence des patterns de performance atypiques, suggérant des

---

<sup>16</sup> Recouvrant dans une large part la dichotomie de Kissine (2012) qui oppose les aspects intersubjectifs ou non-intersubjectifs de l'utilisation du langage

mécanismes alternatifs par rapport à des enfants à développement typique (Marocchini et al., 2022). Similairement, Andrés-Roqueta et Katsos (2020) observent que des enfants autistes comprennent aussi bien des implicatures scalaires que des enfants neurotypiques appariés en niveau de langage (avec une tâche de jugement binaire) mais qu'ils réussissent moins bien certains items des Strange Stories (Happé, 1994) impliquant du faire-semblant, une blague, un pieu mensonge, de l'ironie et une métaphore, utilisés comme mesure de pragmatique sociale. De plus, les performances des enfants dans la tâche d'implicatures scalaires étaient associées à leurs performances en langage formel, tandis que les performances aux Strange Stories étaient associées à la fois au langage formel et aux performances dans des tâches classiques de fausses croyances.

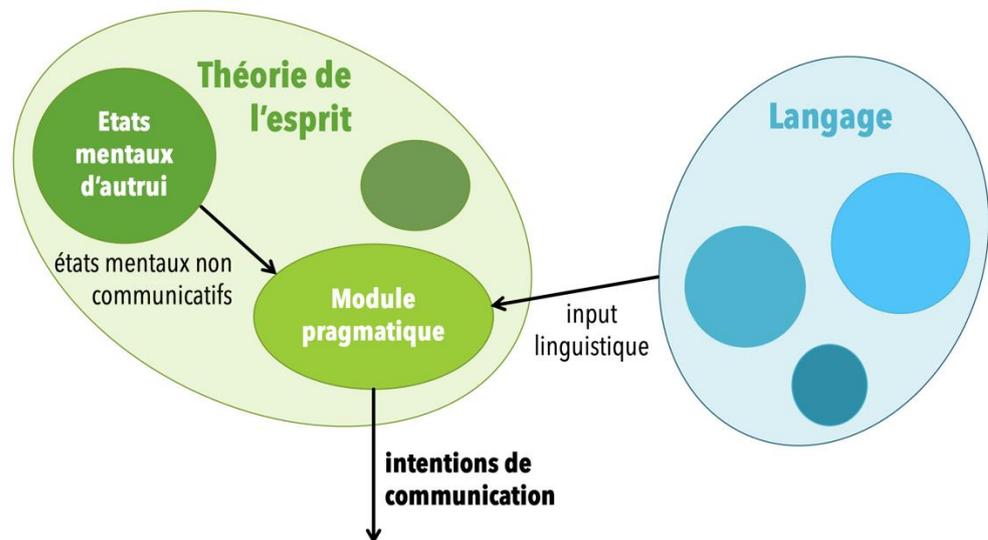
Pour Kissine (2021) ces résultats démontrent qu'un certain nombre de phénomènes pragmatiques (ici entendus comme utilisation contexte-dépendante du langage) comme les implicatures scalaires peuvent être utilisés indépendamment d'un raisonnement sur les états mentaux d'autrui, mettant en doute des propositions telles que la théorie de la pertinence, selon laquelle l'utilisation du langage est fondamentalement une activité de théorie de l'esprit (*chapitre 1*). S'il remet en question la *nécessité* de la théorie de l'esprit pour l'utilisation du langage, il admet toutefois la possibilité que cette habileté *puisse* être mobilisée pour rendre la communication plus efficace, voire qu'elle puisse être « *souvent requise pour une communication réussie* » (p. 7, traduction personnelle). Mais l'absence d'une implication intrinsèque de la théorie de l'esprit pour utiliser un certain nombre de phénomènes pragmatiques (notamment dans des tâches en laboratoire) les rendrait accessibles aux personnes autistes qui, « *quand [ils] utilisent et interprètent le langage en contexte, [...] le font sans se projeter dans l'esprit de leurs partenaires conversationnels* » (p. 7, traduction personnelle)<sup>17</sup>. Quelques précautions semblent toutefois nécessaires avant d'adopter cette perspective. En effet, on peut raisonnablement émettre des réserves quant à la validité de plusieurs prémisses de ce raisonnement. Premièrement, le contraste empirique sur lequel s'appuie cette argumentation demande encore des confirmations et des précisions. Comme détaillé dans les sections précédentes de ce chapitre, le caractère typique du traitement des implicatures scalaires et atypique du langage figuratif dans l'autisme n'est pas encore consensuel dans la littérature scientifique.

Deuxièmement, d'un point de vue théorique, ce raisonnement sous-entend que la théorie de l'esprit constituerait également un bloc monolithique (ce qui est largement contesté, cf. *encart 1*); Kissine (2021), et Andrés-Roqueta et Katsos (2017) argumentent ainsi à l'encontre d'une implication systématique d'un tel processus général dans la compréhension du langage. Cependant, comme évoqué au *chapitre 1*, les développements de la théorie de la pertinence proposent que la compréhension

---

<sup>17</sup> Kissine étend ensuite son argumentation sur le développement du langage en lui-même mais cela se situe hors des considérations de cette thèse.

pragmatique soit prise en charge par un *sous-module* de la théorie de l'esprit spécialisé dans les intentions de communication. La proposition d'un sous-module induit à la fois une interconnexion et une relative autonomie entre les différentes composantes de la théorie de l'esprit (Sperber, 2005). Le sous-module de compréhension pourrait alors exploiter des informations issues d'autres sous-modules, comme ceux renseignant sur les états mentaux (non-communicatifs) d'une personne, pour attribuer des intentions de communication aux locuteurs, sans que ces informations ne soient *nécessaires* au fonctionnement du sous-module pragmatique (Mazzarella & Noveck, 2021). Pour rappel, ce sous-module établirait l'intention de communication déterminant un énoncé en suivant la voie du moindre effort, jusqu'à atteindre les attentes de pertinence optimale de l'auditeur. Ce processus est en lui-même un raisonnement sur les états mentaux d'autrui, mais les attentes de pertinences pourraient elles-mêmes être plus ou moins sophistiquées, selon la précision d'autres sous-modules (non-communicatifs). Mazzarella et Noveck (2021, p. 11) contre-argumentent ainsi qu'« *alors que l'interprétation du langage repose toujours sur la même procédure inférentielle, centrale, ses opérations peuvent être affectée par la qualité des informations qui nourrissent le module de compréhension* » (traduction personnelle ; voir une représentation schématique en figure 3-10). Les effets développementaux pourraient ainsi être expliqués par l'intégration par le module pragmatique d'informations de plus en plus précises issues d'autres sous-modules de la théorie de l'esprit (Sperber, 1994). Cette position clarifierait également le sens à donner à ce qu'Andrés-Roqueta et Katsos (2017) nomment 'normes pragmatiques' et qu'ils décrivent comme nécessaires aux différents sous-types de procédures pragmatiques (linguistiques ou sociales) qu'ils exposent, et pourrait correspondre à ce sous-module pragmatique.



**Figure 3-10 :** Illustration du processus d'interprétation du langage d'après la théorie de la pertinence (Sperber, 2005; Sperber & Wilson, 1996), tel que décrit par Mazzarella et Noveck (2021)

Troisièmement, un tel raisonnement utilise l'autisme comme une condition expérimentale caractérisée par une absence de théorie de l'esprit, comme en témoigne la citation de la page précédente. Or, nous avons également vu au début de ce chapitre

que ce modèle semble à ce jour largement réducteur, compte-tenu de l'état actuel des connaissances : les personnes autistes ont des compétences en théorie de l'esprit, bien que probablement différentes, et leurs difficultés se manifestent généralement avec du matériel ou des situations neurotypiques, plutôt que dans l'absolu (Marocchini, 2023). Les difficultés d'interaction (et de compréhension pragmatique mutuelle) dans des interactions naturelles entre personnes autistes et non-autistes pourraient donc également être attribuées à un défaut d'ajustement des partenaires neurotypiques (e.g., Crompton et al., 2020).

Les données sur l'autisme interrogent donc les modèles de l'autisme, mais également les modèles pragmatiques existant. Il semble toutefois qu'elles ne permettent pas (encore) de déterminer si les inférences pragmatiques reposent sur *une compétence unique ou des processus multiples*. Pour l'heure, les preuves disponibles découragent toute implicature scalaire sur cette expression et promeuvent plutôt une interprétation inclusive de cette disjonction, tant la conjugaison d'une compétence unique engagée dans des processus multiples apparaît également plausible.

## Problématique de la thèse

Lors de cette revue de littérature, nous avons vu que la communication humaine ne peut s'expliquer comme l'utilisation d'un code, mais comme la manipulation de signes pour manifester des intentions de communication ou inférer celles d'autrui. Des inférences pragmatiques sont nécessaires pour interpréter ce qu'un locuteur *veut dire*, toujours sous-déterminé par ce qu'il *dit*. La distinction entre ces deux niveaux de signification a particulièrement été étudiée dans le cas des métaphores et des implicatures scalaires qui sont au cœur de ce travail de thèse. Par rapport à la compréhension strictement littérale des énoncés, inférer les intentions de communication qui les déterminent a donc une importance fondamentale dans le fonctionnement quotidien de l'être humain. De nombreux travaux montrent que cette opération est réalisée au prix d'un certain effort cognitif, qui peut être observé chez l'adulte et qui est également manifeste dans le délai développemental observé dans la mise en place de ces habiletés, ainsi que dans les difficultés dont certaines populations cliniques peuvent faire l'expérience. La compréhension dite pragmatique est en effet typiquement décrite comme se développant tardivement par rapport aux aspects formels du langage, jusqu'à l'adolescence. Toutefois, la recherche s'est récemment plutôt focalisée sur le développement précoce de ces habiletés, ancré dans les compétences langagières et sociales précoces de l'enfant, négligeant relativement leur développement ultérieur, notamment à l'âge scolaire (6-11 ans). Un certain nombre de problématiques demeurent cependant pertinentes aux vues de la littérature. Le développement d'un nouvel outil d'évaluation clinique (le TIPi), offre l'opportunité d'investiguer ces questions.

Pour ce qui est des implicatures scalaires, l'effet développemental robustement observé avec des tâches de jugements binaires (Katsos et al., 2016; Noveck, 2001) pourrait refléter un délai dans la mise en place des opérations pragmatiques, en raison d'une tolérance pragmatique vis-à-vis d'énoncés sous-informatifs (Katsos & Bishop, 2011), plutôt que d'une immaturité dans le développement pragmatique lui-même. N'étant qu'une mesure indirecte des implicatures scalaires, les tâches de jugements binaires pourraient ainsi avoir biaisé les résultats classiquement observés. Dans ce contexte, nous proposerons dans une *première étude* de rechercher cet effet développemental classique avec une nouvelle tâche évaluant *directement* l'interprétation de termes scalaires chez les enfants, adaptée de Pouscoulous et al. (2007, exp 2). Alors que les études précédentes comparaient généralement des groupes d'âges aux effectifs relativement réduits, et aux frontières arbitraire, l'inclusion d'un large échantillon d'enfants d'âge scolaire permettrait en outre de préciser la trajectoire développementale de ces implicatures en fonction de l'âge, de façon continue. Enfin, l'adaptation que nous proposons permet une mesure des temps de réponse, mesure

typiquement plus difficile à mettre en place auprès d'enfants. Ce type de variable s'étant révélé informatif chez l'adulte, son application chez l'enfant, pour la première fois à notre connaissance, devrait par ailleurs permettre d'éclairer les mécanismes qui déterminent les comportements des enfants face à ces implicatures. En particulier, cela devrait permettre d'établir si les réponses logiques, par rapport à des réponses pragmatiques, reflètent une immaturité des compétences pragmatiques ou plutôt une tolérance pragmatique vis-à-vis d'énoncés sous-informatifs.

Pour ce qui est des métaphores, des travaux récents ont suggéré que les effets développementaux tardifs qui sont classiquement observés (e.g. Winner et al., 1976) pourraient être attribués à la forte implication des habiletés verbales et métalinguistiques dans les paradigmes expérimentaux qui sont utilisés, plutôt qu'à la compréhension pragmatique *per se* (e.g. Pouscoulous, 2014). Les études récentes continuant de montrer un développement de la compréhension des métaphores à l'âge scolaire s'appuient en effet sur des tâches d'explication verbale (Lecce et al., 2019) ou de manipulations métalinguistiques (e.g. Deckert et al., 2019). Nous proposerons donc dans une *deuxième étude* une nouvelle tâche de référence métaphoriques sur la base de celles de Noveck et al. (2001) et de Van Herwegen et al. (2013), qui limite et contrôle l'implication des habiletés verbales et métalinguistiques et qui permet également de récolter les temps de réponse, afin d'examiner si de telles conditions confirment que l'âge scolaire est une période clef dans le développement des métaphores. Par ailleurs, les prédicteurs cognitifs de ces habiletés font encore l'objet de débats. Les compétences langagières formelles et la théorie de l'esprit sont deux candidats privilégiés, notamment suite aux travaux réalisés sur l'autisme, mais ils ont peu été évalués de façon conjointe dans un même échantillon. De plus, la plupart des tâches de théorie de l'esprit mobilisées jusqu'alors sont fortement verbales, rendant difficile d'observer la contribution respective de ces deux habiletés. Nous associerons donc à cette nouvelle tâche de références métaphoriques une tâche de compréhension de phrases et une tâche non-verbale de théorie de l'esprit, issue de celle de Langdon et Coltheart (1999) que nous avons adaptée sur tablette et validée pour une utilisation auprès d'enfants (étude supplémentaire en *annexe B*). Ce dispositif nous permettra d'observer l'association entre la compréhension des métaphores et, respectivement, le langage et la théorie de l'esprit, et leur évolution au cours du développement. Les travaux de Lecce et al. (2019) suggèrent en effet que l'influence de la théorie de l'esprit diminuerait avec l'âge. Ces données devraient également permettre de positionner cette nouvelle tâche dans la dichotomie introduite par Andrés-Roqueta et Katsos (2017) entre pragmatique linguistique et pragmatique sociale.

Finalement, ces deux nouvelles tâches pourront être mobilisées pour examiner le développement autistique des inférences pragmatiques dans une *troisième étude* rapportant des données cliniques préliminaires (voir aussi l'article en *Annexe D*). En effet, nous avons vu qu'en dépit des représentations caricaturales associées à l'autisme, un contraste semble se dessiner entre des phénomènes pragmatiques qui seraient atypiques, notamment les métaphores (Morsanyi et al., 2020), et d'autres typiques,

comme les implicatures scalaires (e.g. Chevallier et al., 2010). Toutefois, ce contraste a peu été observé au sein d'un même échantillon de personnes autistes, et chacune des lignes de recherche présente des limitations qui appellent à la prudence. Pour ce qui est des métaphores, comme pour le développement typique, les difficultés classiquement associées à l'autisme ont été attribuées à des difficultés verbales plutôt qu'à des caractéristiques proprement autistiques (e.g. Kalandadze et al., 2019). Pour ce qui est des implicatures scalaires, les performances typiques de participants autistes sont établies avec des tâches de jugements binaires, mais d'autres types de tâches semblent révéler des différences associées à l'autisme, qui demandent confirmation (e.g. Mazzaggio et al., 2021). En utilisant les tâches expérimentales exploitant les réponses comme les temps de réponses des participant, et validées dans les deux études précédentes, nous proposons ainsi d'évaluer (i) si des enfants autistes se comportent de façon typique face à des énoncés sous-informatifs comparables à ceux utilisés dans les tâches de jugement binaires, mais dans une tâche expérimentale évaluant *directement* ces implicatures scalaires, avec un mode de réponse non-verbal ; (ii) si des enfants autistes réalisent des performances plus faibles qu'attendu dans une tâche de compréhension de métaphores qui limite et contrôle la demande verbale et métalinguistique ; et enfin (iii) si ce contraste entre implicatures scalaires typiques et métaphores atypiques peut être observé au sein d'un même échantillon.

# CONTRIBUTIONS EXPÉRIMENTALES

## Chapitre 4 • Développement typique des implicatures scalaires (étude 1)

### Article *en préparation* : Setting *some* tables: an action-based response time paradigm to study the development of scalar implicatures

Nicolas PETIT, Jérôme PRADO, Ira NOVECK

---

#### Abstract

Compared to adults, children are often assumed to struggle with scalar implicatures. Yet, this was mostly observed with indirect measures that are sentence verification tasks, and this developmental effect could be explained by an *emerging pragmatic capacity* as well as by a pragmatic competence available early on, but masked by experimental features (*pragmatic tolerance*). Here, an action-based tablet paradigm that directly measure children's scalars interpretations and collects response times was used with 249 typically developing children, between 6 and 11 years old. Results confirmed at a large scale the developmental effect and revealed that pragmatic interpretations of scalars were slower than semantic interpretations and control responses, in 6-7 years-old children. Thus, semantic responses appeared immature compared to pragmatic responses, which plays in favor of a late *emerging pragmatic capacity*. This suggests that children can use informativeness when asked to, but that they struggle to do so spontaneously to infer speakers' meaning.

---

Upon hearing a speaker say, “*Mary ate some cookies*”, one would be entitled to conclude that *Mary did not eat all of the cookies*. This is remarkable because – semantically speaking – the meaning of *some* is compatible with *all*. This illustrates how enriching the utterance to mean *Mary ate some but not all of the cookies* results, not from a logical inference but, from a pragmatic enrichment. This inference is usually described with respect to Grice's (1989) model of human communication, according to which a speaker's intended meaning can be derived from the assumption that the speaker is cooperative

and respects a set of conversational maxims. In this case, the *maxim of quantity* is invoked to enjoin interlocutors to be as informative as possible. The upshot is that a hearer can reason that the speaker's choice of using the quantifier *some* – instead of the more informative *all* – implies that Mary ate *some but not all* the cookies. In other words, a less informative term, like *some*, indicates that a more informative term from the same scale is not the case (see Horn, 1989). This is why the inference is usually referred to as a *scalar implicature* which can be readily extended to other scales, e.g., the use of *might* implies the exclusion of *must* and the

disjunction *or* can be taken to mean that the conjunction *and* can be eliminated from consideration (which provides for the exclusive reading of disjunctions).

Scalar implicatures are exemplary of phenomena that are investigated in linguistic-pragmatics, which is an area concerned with the way the intended meaning of an utterance is derived from its components. With the advent of Experimental Pragmatics (see Noveck, 2018; Noveck & Sperber, 2004), scholars have been in a better position to evaluate how well theories account for pragmatic phenomena as they are captured in laboratory experiments. One of the established findings in the experimental scalar implicature literature (starting from Noveck, 2001) is that children are less likely than adults to draw out pragmatic readings, even if they appear competent in evaluating a sentence's semantic meaning. When presented with a sentence like "*Mary ate some cookies*", multiple studies indicate that a child would be more likely than an adult to accept that sentence as true even when it was shown that Mary had eaten *all* of the cookies (Barner et al., 2011; Chierchia et al., 2004; Guasti et al., 2005; Katsos et al., 2016; Katsos & Bishop, 2011; Papafragou & Musolino, 2003; Pouscoulous et al., 2007).

In what follows, we take the following three steps. First, we describe in greater detail the experimental background on scalar implicatures, which will focus on the developmental effects but will also describe relevant findings from the adult literature. This summary will allow us to describe the range of materials and paradigms that have been used in the study of scalar implicature as well as to underline what is missing from the developmental literature (viz. reaction time measures). Second, we will introduce two of the most prominent accounts for the developmental effects – one that we will call the *emerging pragmatic capacity*

account, which assumes that pragmatic competence evolves with age and subsequent to semantic competence and another, known as *pragmatic tolerance*, which assumes that pragmatic competence is available to all from a young age but that it is masked by features of experimental designs. Finally, we will present our own paradigm which includes reaction time measures. These sorts of measures will put us in a stronger position to address the debate between the two contrasted accounts.

### **Background on the development of scalar implicatures**

The developmental effect related to scalar implicatures was first reported by Noveck (2001). To make it concrete, consider Experiment 3's double-blind study, in which participants were presented out-of-the-blue underinformative sentences such as "*Some giraffes have long necks*" along with multiple control items. The task required participants to determine whether or not they agreed with each statement (by responding *Yes* or *No*, respectively). With respect to this example, which can be considered infelicitous (since all giraffes have long necks), there are two readings. With a semantic (*some-and-possibly-all*) reading of the quantifier, one should expect a *Yes* response and with a pragmatic (*some-but-not-all*) reading, one could expect a *No* response. Noveck (2001) reported that adults were equivocal by responding *No* 59% of the time. That overall rate reflected a bimodal distribution: most adults were either consistently pragmatic or consistently semantic in their responses. In contrast, two groups of children (who were 7-8 years old and 10-11 years old) largely accepted these underinformative sentences by responding *Yes* (in 89% and 85% of cases, respectively). Importantly, all three groups performed comparably on control items which included True and False items

using *all* (*All robins have wings; All dogs have spots*) as well as factually true and patently false items using *Some* (*Some birds live in cages; Some stores are made of bubbles*).

Since Noveck's (2001) paper, this developmental trend has been investigated and replicated across many paradigms (e.g., Barner et al., 2011; Chierchia et al., 2004; Guasti et al., 2005; Katsos et al., 2016; Katsos & Bishop, 2011; Papafragou & Musolino, 2003; Pouscoulous et al., 2007). Typically, test items rely on general knowledge (as in Noveck, 2001, Experiment 3 above) or on visually presented context that can be experimentally manipulated (see Noveck, 2001, Experiment 1). To illustrate how one can manipulate a visual scene, consider Papafragou & Musolino (2003) who presented five-year-old participants a scenario with horses, all of which were shown to jump over a fence; afterward, the experimenter (in the guise of a puppet) produced the sentence "Some of the horses jumped over the fence." At this point, the young children were presented a metalinguistic question, viz. had the puppet "answered well." Nearly all adults replied negatively with respect to the underinformative sentences whereas most 5-years-olds accepted them; as expected, participants performed comparably on control sentences.

The Papafragou & Musolino (2003) paper was notable because it augured in a series of studies that were designed to make scalar implicatures more accessible to children. Even though the authors' experimental efforts increased rates of pragmatic responding to underinformative sentences among the young children, the developmental trend persisted, with adults rejecting the underinformative items almost categorically. Guasti et al. (2005) also showed that training among 7-year-olds can encourage pragmatic responding that

resembled adults' while using the materials from Noveck's (2001) Experiment 3. Their training consisted of making the goals of answering more explicit, by showing, for example, a grape to young participants and asking whether it was better to describe it as a grape or as a fruit. The 7-year-olds who received such training later appeared adult-like when asked to carry out the task in Noveck's (2001) Experiment 3; those who did not receive training did not answer in an adult-like way. Interestingly, the training no longer had salutary effects on the trained children when they were revisited just one week later. Pouscoulous et al. (2007), whose paradigm will be a model for our experiment, also endeavored to make scalar implicatures more accessible to children. However, they did not use the infelicitous out-of-the-blue items (like in Noveck's Experiment 3) and they did not rely on training. Instead, they modified the original paradigm to remove metalinguistic costs in the response format and make it action-based. They presented participants with a set of five boxes before asking participants to comply with a puppet's wish, e.g., "*I would like all the boxes to contain a token.*" Importantly, there were three different starting points before each expressed wish: i) the case in which there was a token in each of the boxes, ii) the case in which a token is in none of the boxes or; iii) the case in which there is a token in only some of the boxes. Participants were put into a situation where they could change the setting if they thought it necessary. Different quantifiers were used in the wish, viz. *all*, *some*, *none* and *some are not* (as in "*I would like some boxes to not contain a token*"). The critical condition arises when a token is initially placed in each box and the puppet's wish is: "*I would like some boxes to contain a token.*" A participant relying on the semantic reading of *some* would not need to make any change to this configuration, but a participant drawing the scalar inference would be expected to remove at least one of

the tokens. While the paradigm encouraged pragmatic inferencing generally, the authors still reported a developmental progression. To wit, Pouscoulous et al. showed that 32% of 4- to 5-year-old children's responses were "logical" (i.e. they did not make any changes to the scenario) and that this percentage decreased with age. To put it another way, pragmatic responding increased with age. Miller et al. (Miller et al., 2005) had also demonstrated that action-based paradigms could facilitate children's scalar inferences in a drawing task. In one of their critical items, for example, they presented participants with four mouth-free faces, and asked them to "use the crayon and make some faces happy" (p. 5). In this condition, half of the productions of 4 to 5 years old children was considered pragmatic (that is, they did not draw smiles on *all* of the faces), contra 96% of adults. The authors argued against Noveck's claim by saying that children were able to compute scalar inference, but the developmental effect remained evident<sup>18</sup>.

Eye-tracking techniques (Huang & Snedeker, 2009) have also been employed to investigate the development of scalar implicatures. This more subtle technique has participants to listen to sentences such as "Point the girl that has some of the socks." The cleverness of the paradigm rests on providing a scene in which there are four human figures, two girls and two boys, each adjacent to objects (or to nothing). Given this sentence along with a scenario in which one of the two girls had two socks next to her (along with a boy who also had two), one might think that participants' looks would fall quickly on the girl who had *some but not all* of the socks. On the other hand, if the word *some* is not readily enriched and

the other distributed objects were the initially-homophonic *soccer balls* (such that one girl had all of the scenario's soccer balls) then it would be reasonable for the listener to wait until the end of the sentence in order to disambiguate which object the speaker was talking about. Adult participants' eye movements indicated that from the moment they hear the word *some* they anticipate what picture needs to be pointed at, demonstrating that they interpreted *some* as *some but not all* on line. Five years-olds, on the contrary, did not show such anticipation and looked to the target only after receiving the disambiguating final word. Note that this means that they considered the girl with *all* the soccer balls a potential referent while hearing *Point to the girl with some of the*. It could be argued, though, that children have only a limited time to enrich scalar terms in this paradigm (the time between "some" and the disambiguating phoneme is uttered) and that they might have shown evidence of pragmatic inference if they had had more time. That said, children did not seem to experience major slowdowns among the control items. Taking all the data together, it appears to point to children being more likely to entertain the semantic reading when compared to adults.

### Related (adult) studies using reaction times

Before we present two opposing accounts of this developmental-pragmatic effect, it behooves us to summarize findings from adults that use response times, partly because the patterns reported in this part of the literature resonate with those concerning the developmental progression. We will provide two seminal examples. Consider first Bott & Noveck (2004) who asked participants to evaluate out-of-the-

---

<sup>18</sup> In another condition they showed that putting a contrastive accent on the scalar term (i.e. "use the crayon to make *SOME* faces happy") also

helped children to derive the inference, but again, the developmental effect remained.

blue categorical sentences (e.g. *All robins are birds*), including underinformative ones such *Some cats are mammals*. Given that true responses to the underinformative items indicate a semantic (*some and perhaps all*) reading of such sentences and that false responses point to pragmatic (*some but not all*) readings, it becomes crucial to know which is faster. If the developmental findings can be generalized, one would expect that those who give pragmatic responses to take longer than those who give responses that point to semantic readings. This is precisely what was reported over four experiments, including one (Bott & Noveck, 2004, Experiment 4), which showed that short imposed time limits (giving just 900 msec to respond instead of 3000 msec) increased logical responding (i.e. it lowered the likelihood of finding pragmatic responses among participants). Importantly, when participants provided responses to underinformative utterances that were consistent with semantic readings, their speed of responding was typically comparable to the control sentences. This was not the case for those responses that were consistent with pragmatic responding. Ronderos and Noveck (2023) recently replicated this finding in an online version of this task, and provided explanations for that slowdown. They argued that the temporal cost associated to implicature derivation is, at least partly, a consequence of how participants think about informative intentions of the speaker. To support their claim, they provided evidence that response times to underinformative items reduced over the course of their experiment, ending with comparable response times than control items, which is not predicted by general mechanisms that could underlie implicature resolution. In a subsequent experiment, they showed that changing the speaker during the experiment generated slowdowns in response times of underinformative statements, showing

participants need to adjust to another's mental states.

Another critical study using reaction times came from Breheny et al. (2006), who asked participants to read stories in a self-paced fashion. Stories were prepared so that contexts encouraged either semantic interpretations of scalar expressions or pragmatic ones. Their results provided evidence showing that scalar implicatures are not only context-dependent but that the very same set of words can require more effortful processing when the context encourages a pragmatic, as opposed to a semantic, reading.

Here, we point out that reaction time (or reading time) experiments have not been employed among children and for obvious reasons, i.e. it is challenging to investigate children's performance with subtle reaction time measures especially given that their skills are likely to vary widely at each age. This is unfortunate for the literature because it would be a useful technique to add to this literature. This is one important challenge that we take up in our work.

### **How does one account for the developmental effect?**

While there is a consensus that developmental progressions appear to trend toward pragmatic readings with age, there is less consensus about how to account for it. Here, we will compare two possibilities. The first is that, given the effort involved in producing pragmatic readings, children are just less capable than adults in deriving scalar implicatures. We will refer to this hypothesis as the *emerging pragmatic capacity* account. In other words, the above results can best be described as reflecting a delay in applying pragmatic operations to utterances (see Noveck, 2001; Guasti et al., 2005); in light of the delay, semantic readings are more than adequate. The second alternative is that children have a

pragmatic competence early on in their cognitive development but that their threshold for applying a pragmatic reading is more flexible. That is, they notice an infelicity when one arises but they are more tolerant of it before invoking pragmatic operations on their interpretation. We will refer to this approach, as does the literature, as the *pragmatic tolerance* account, whose main advocates have been Katsos & Bishop (2011).

Given its nuance and originality, let us briefly review the pragmatic tolerance account. Katsos and Bishop (2011) formulated both a methodological criticism of the existing evidence and a theoretical claim. Methodologically, they argued that judgment tasks should not be *directly* assimilated to a measure of scalar inference (see also Katsos, 2014). An underinformative utterance (such as *Some elephants have trunks*) is both logically true and sub-optimal regarding informativity. Having to judge whether such an utterance is true or false thus forces the participant to *pick a side* but the answer given does not necessarily reflect whether they detected the ambiguity, or how they would have spontaneously interpreted this term in the real world, but rather a response strategy. This methodological criticism leads Katsos and Bishop to a theoretical possibility: children might be able to perceive that an utterance is not informative enough, but not consider it as reason enough to reject it, as adults often do. To support their claim, they first replicated the classical finding with a judgment task in which they asked participants to judge whether sentences uttered by an English-learning character were *right* or *wrong*. As expected, 5-to-6-years-old children tended to accept underinformative sentences with *some* (74% of acceptance) while adults generally rejected them. Then, with the same material, participants “*were asked to reward his response using a 3-point scale*

*consisting of different-sized strawberries.*” With this set of response options, children and adults typically used the smallest and largest strawberry to correctly characterize the semantically false and true sentences, respectively, while both groups used the intermediate one to judge underinformative sentences. In other words, it showed that children can be as sensitive as adults to the informativity of an utterance. That is, they have the pragmatic capacity to judge the felicity of an utterance, but are more *tolerant* with violation in standard judgement tasks. Part of the authors’ motivation is to bring the developmental scalar implicature in line with other developmental pragmatic phenomena, notably intention attribution and informativity management, which are present very early in child development (see Stephens & Matthews, 2014, for a review of early pragmatic development). Plainly, the idea that children could use pragmatic reasoning to learn new words but to not derive scalar implicature at school-age is puzzling without some kind of pragmatic module that allows for detecting infelicity.

### **Filling a gap**

The current study uses a novel paradigm inspired by Pouscoulous et al. (2007) which puts us in a position to complement the literature on three different aspects. One is that it extends and refines the age of participants by covering the whole range of 6 to 11 years old (Pouscoulous et al. used 3 different age groups, viz. 4-year-olds, 5-year-olds and 7-year-olds plus adults). It thus offers a *continuous* indication of pragmatic performance with scalars among school-age children. Second, whereas it has been argued that judgement tasks only provide indirect evidence of scalar inferences derivation, ours directly reveals how scalars are interpreted. The third is that it provides added value by including a new dependent variable for children -- reaction

time data. This is introduced in order to better characterize young children's growing appreciation for pragmatic readings. This was accomplished by shifting the paradigm to tablets.

The current method allows us to compare children's semantic vs pragmatic interpretations of *some*, which is of particular interest for the questions raised by the debate above. That is, if children are sensitive to informativeness violation but are tolerant of it, both semantic and pragmatic interpretations of *some* should be more effortful (i.e. longer) than control items. Indeed, sensitivity to informativeness sheds light on the ambiguous nature of underinformative statements, which must be solved to infer the speaker's intention, in any case (no matter what interpretation is made). But if children simply become more pragmatic with age, semantic responses should reflect an absence of pragmatic reasoning and, as such, be a) faster than pragmatic readings and b) as fast as control items. In consideration of the extant data on scalar inference-making among both children and adults, we predicted the latter.

## Method

### Population

Two hundred and forty-nine children, ranging from 6- to 11 years old (mean = 8.4, SD = 1.4) and who were homogeneously distributed in this age range (see precise distribution by class in table 1), participated in this study. All the children regularly attended an elementary school (*école primaire*) and were native speakers of French (though bilingualism was not disqualifying). Parents provided informed consent. The children's participation followed the reception of parental reports indicating that there were no

neurodevelopmental disorder (learning, autism spectrum, attention-deficit, or language-related disorder), and no sensory (visual or auditory) or motor disorders that would prevent the child from using a tablet.

The children were recruited from two French schools, one urban private school in an economically privileged area (School A) and, to provide social balance, one rural public school situated in a less economically privileged region (School B). The children's characteristics, including those provided by the children's parents, are detailed in Table 4-1. These include an estimation of family income through the Family Affluence Scale (FAS, Boyce et al., 2006), which is based on relatively non-intrusive questions and provides a continuous measure ranging from 0 to 9 that has proven to be a good indicator of family wealth across countries (Boyce et al., 2006; Currie et al., 2008). Questionnaires also asked parents about their education levels, which led to a code ranging from 0 (no diploma) to 7 (PhD); parents' mean education level is thus reported for each child. When data from only one parent is available (6 % of children), the score of that one parent is reported. These measures confirmed that parental education levels and income as indexed by the FAS were clearly higher in School A than in School B ( $t(234.41) = 9.95, t < .001$  and  $t(243.83) = 5.70, p < .001$ , respectively). The FAS and the Parental Educational Level indicators were combined to provide a composite socio-economic status (SES) result varying from 0 to 100 %, with each indicator providing half of the composite. Data were missing for 2 children and replaced by group mean for the analysis.

This study received ethical approval from the Comité de Protection des Personnes Sud-Est I (ID RCB 2019-A01721-56,) and received authorizations to conduct the research from the two schools.

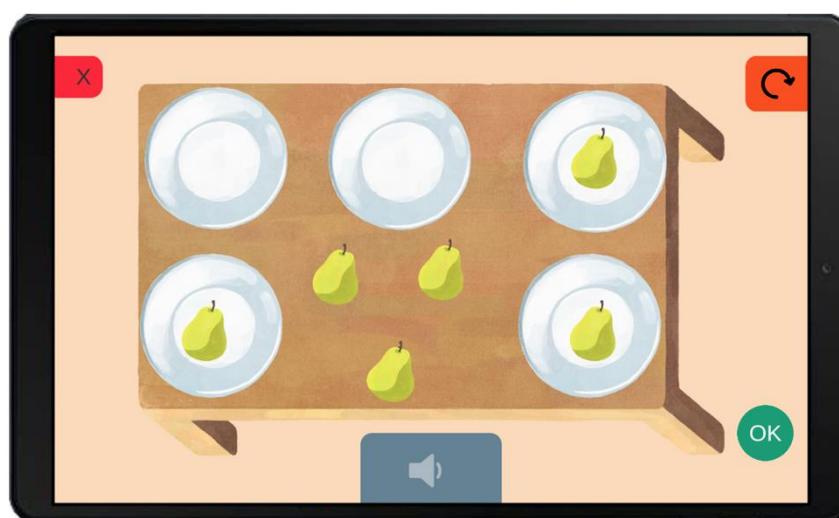
**Table 4-1:** Sample characteristics by class

	1 <sup>st</sup> grade (CP)	2 <sup>nd</sup> grade (CE1)	3 <sup>rd</sup> grade (CE2)	4 <sup>th</sup> grade (CM1)	5 <sup>th</sup> grade (CM2)	Whole group	Class difference
N	52	58	50	42	48	249	/
Proportion girls	48%	52%	54%	46%	44%	51%	No ( $\chi^2(4) = 1.25, p = .9$ )
Mean age in years (SD)	6.5 (0.3)	7.5 (0.3)	8.4 (0.4)	9.5 (0.3)	10.4 (0.3)	8.4 (1.4)	Yes ( $F(4,245) = 1126, p < 10^{-15}$ )
Mean FAS (SD)	5,7 (1,2)	5,7 (1,5)	5,6 (1,5)	6,1 (1,5)	5,6 (1,7)	5,7 (1,5)	No ( $F(4,243) = 1.14, p = .34$ )
Mean parental education level (SD)	4,2 (1,6)	4,5 (1,9)	4,2 (1,9)	4,3 (1,8)	4,0 (1,7)	4,2 (1,8)	No ( $F(4,243) = .53, p = .71$ )
Mean SES Index (SD)	57 (14)	59 (18)	57 (17)	61 (16)	56 (16)	58 (16)	No ( $F(4,242) = .7, p = .59$ )
Proportion from school A	46%	57%	40%	57%	53%	51%	No ( $\chi^2(4) = 4.42, p = .35$ )

## Material

Scalar implicatures are assessed with the TIPi (*Test des Inférences Pragmatiques informatisé*), a new clinical tablet-based test developed to assess a variety of pragmatic inferences in children. The scalar task is inspired by Pouscoulous et al. (2007, Experiment 2). In the current tablet-based version of that task, each trial begins with the child being presented with a screen displaying a table with 5 plates and 6 fruits. Prerecorded audio instructions were prepared so that the children would receive information that would allow them to help set the table, according to the speaker's wishes. For example, consider the scenario below which displays 6 pears (Figure 4-1).

Upon being shown the screen, the child clicks on a speaker icon which provides the speaker's wish concerning the setting of the table. For example, for the above figure, the child hears "I want all plates to have a pear" (*Je veux que toutes les assiettes aient une poire*). This is a case where the child can move two of the unaffected pears to the remaining empty plates so as to satisfy the wish. Once the request is satisfied, participants can then validate the trial by clicking on "OK". Note, however, that the display could also begin with each plate already having a pear, in which case the child would have nothing further to do other than validate the scene. The child is allowed to press a repeat button to hear the speaker's wish again and for as many times as desired.



**Figure 4-1:** Initial display of an item of the task

The task used three different quantifiers in the instruction, namely *all*, as in the example (*tous* in French), *some* (*certain*) and *none* (*aucun*). This accounts for the quantifier variable. As far as the distribution of fruit goes, a test scenario starts in one of three ways (visual contexts): 1) where each plate is empty (thus, the trial begins with all six exemplars of the fruit appearing in the center of the table); 2) where each plate contains a fruit (this means the one extra fruit remains unplated, in the center), or 3) where only some of the plates (that is, 2 or 3) have a fruit; this means the remaining fruit are seen in the center of the table. Combining these two variables leads to 9 possible conditions. The critical condition arises when all of the plates contain, say, an apple, and the instruction is “I would like some of the plates to have an apple”. Here, a semantic reading of *some* (which is compatible with *all*) should lead the child to click on OK. However, if children enrich *some* to mean *some but not all*, they are expected to remove at least one fruit before validating it. Video examples of control and critical items are available in online supplementary materials (<https://osf.io/rbgj3/>).

Five different fruits are used (apples, oranges, pears, strawberries and bananas) to limit repetition. Generally, the display

includes six exemplars of the same single fruit (e.g., 6 apples or 6 strawberries). However, to vary the task and to dissimulate its goal, 8 filler items were also included in which different fruits are presented on the screen; in these specific items only, fruit selection is involved in the answer. For example, such an instruction can be “I would like all plates to have a banana” when 5 bananas *and* 1 pear are displayed at the center of the table.

The distribution of items across conditions (Figure 4-2) was made to ensure that the number of underinformative *some* items (N = 7) was large but remained comparable to the amount of control quantifier items (*none* and *all*, N = 8), and that when filler items are included, *some* remains less presented than the other quantifiers taken together (13 presentations of *some* vs 16 presentations of *none* or *all*). Moreover, even if in the critical condition all of the plates must be initially filled (visual context), we made sure that, experimental and filler items taken together, it remained comparable with the other visual contexts (9 items begin with all plates empty, 9 with only some filled and 11 with all plates filled).

The task begins by providing training with 4 supplementary items in order to familiarize the child with the response format and with the possibility of a)

**QUANTIFIER**  
« I want \_\_\_\_\_ of the plates to have a [fruit] »

		NONE	SOME	ALL
<b>VISUAL CONTEXT</b>		1 item +2 fillers	1 item +2 fillers	1 item +2 fillers
		1 item +2 fillers	3 items	1 item +2 fillers
		2 items	<b>7 items</b>	2 items

**Figure 4-2:** Items distribution in the scalar task

removing or adding fruits or else b) directly validating the configuration. Feedback regarding the success is given during the training, and only during the training. The sentence types are different from those used in the test items. Three of these items use exact numbers (for example, “I would like 3 plates to have a banana”), one of which needing to add some fruits, the two others allowing direct validation without any change. Note that we did not include an item in which the number underspecifies the total number of fruit because this would invite a scalar implicature for numerals. Moreover, *incorrect* feedback is given only if the child’s answer is *semantically* incorrect (e.g., if the instruction is “I want 3 plates to have a banana”, leaving 0, 1 or 2 plates would be considered *incorrect* and leaving 3 plates full as well as 4 or 5 plates would be considered *correct*), so as not to bias an individual’s preferences regarding implicatures. The fourth item requests the participant to withdraw all fruits from the plates (“I would like all of the plates to be empty”). The task only began once training was successfully completed.

All instructions were prerecorded by the first author and the consistency of the prosody was checked with Praat software. Responses are automatically recorded, as well as response times, which are measured from the moment a disambiguating phoneme (/t/ in “tous”) in the instruction is presented to the participant to the moment the child presses the *OK* button. The mean duration from the beginning of the instruction to disambiguating point was 638 ms (SD = 73.5, range 566-805 ms). Items are pseudorandomized and regularly displayed in a unique list which is used with each participant, according to the following constraints: the 2 first items were not critical items, two critical items were separated by at least 2 control or filler items, and different fruits were used for two consecutive items.

## Procedure

While each child was assigned a tablet and a headset, the children were tested collectively in classrooms during school time. The operation of each device was checked before testing began. The task in the current experiment was part of a larger project that includes a variety of other tasks that are not presented here. The current task took roughly 4 minutes and was proposed after the other tasks had already required roughly 20-25 minutes of each child’s time. In classrooms 15 to 25 children completed the tasks simultaneously at their own pace. Two experimenters remained available in case of problem, but no help regarding the task could be provided after the training was completed.

## Data coding and analysis

In order to calculate error rates, we took the following steps. First, training and filler items were excluded. Second, responses to the quantifier *all* were coded as correct when the final outcome before validation had 5 out of 5 plates filled and incorrect otherwise. Third, responses to *none* were coded as correct when the final outcome before validation had 0 out of 5 plates filled and incorrect otherwise. Responses to *some* were coded as incorrect when the final result had 0 plates filled, and correct otherwise. In the critical condition, responses were furthermore coded as semantic (when the final result had 5 out of five plates filled) or pragmatic (when the final result had 1, 2, 3 or 4 plates filled).

For response times analysis, incorrect answers were excluded, as well as non-critical *some* items which were not relevant for this comparison; indeed, there is no way to differentiate pragmatic versus semantic interpretations of these instructions. The quantifier and answer variables were then combined in a novel variable with three conditions: pragmatic answers vs semantic

answers in the critical condition (Figure 4-2, green), vs correct responses to control quantifiers (Figure 4-2, orange); this new variable will be referred to as *answer-type*. Importantly, response times are expected to take a different time when children move objects compared to when they validate scenarios directly (without making change), since different decisions are involved. We thus included this as a binary control variable (*hasMoved*). Additionally, when children moved fruits before validating, the total length of time per trial should naturally depend on how many fruits were moved. We controlled for this with a continuous control variable (*moves*) capturing the absolute difference between the initial setting (say, 3 plates filled) and the child's response (say, 5 plates filled,  $moves = ABS(3-5) = 2$ ). These two variables are systematically included as covariates in our models. This is meant to allow the comparison of pragmatic responses in the underinformative 'some' items (cases where participants made changes), to semantic responses in the underinformative 'some' items (cases where no change is required), and to control items (where both situations happened).

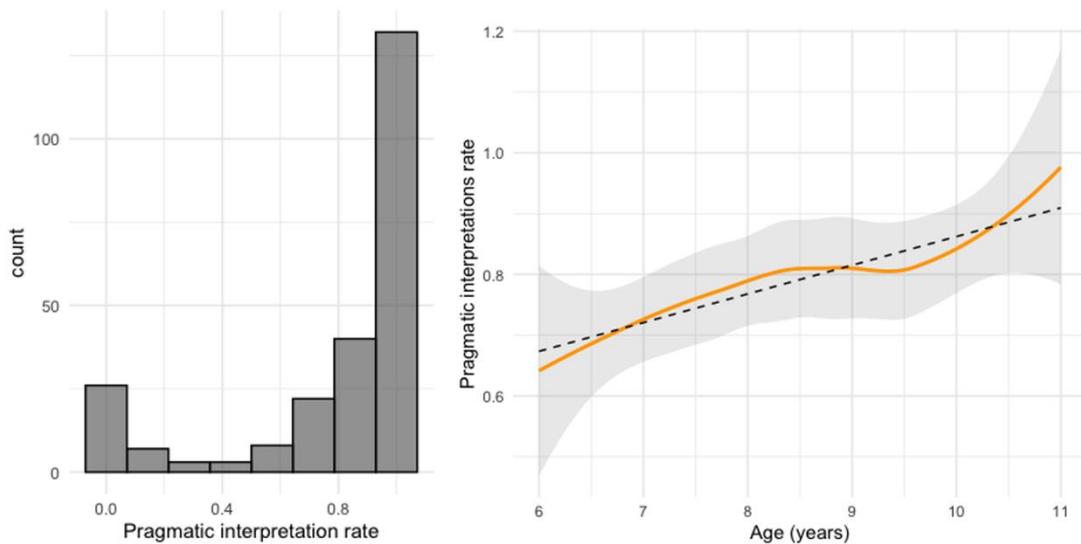
Data are analyzed with generalized (for performance) of linear (for RT) mixed-effects models, fitted to the data with the *lme4* package (Bates, Bolker & Walker, 2015) in *R* (R Core Team, 2022). Continuous regressors were centered on the mean. Response times were log-transformed to correct for typical skewness of the distribution. Models' assumptions were checked before reporting their outcome. Fixed effects are assessed with likelihood ratio tests comparing models with and without the parameter in question, using the *anova* function. Post-hoc contrasts were computed with the *emmeans* package (Lenth, 2022) and default single-step method to correct for multiple comparisons.

## Results

### Accuracy

Before considering our findings concerning the underinformative items, we first consider children's performance regarding the semantic content of the sentences used. The results indicated that the task was largely accessible to children, since the overall rate of errors was extremely low (2.2 %). Only 5 participants out of 249 had a 20% or greater error rate (three 1<sup>st</sup> graders and two 2<sup>nd</sup> graders). These five were summarily removed from the sample. When considering the remaining 244 children, the overall error rate was 1.6%. We then fitted a logistic model explaining error with age, SES, quantifier, and position of the item in the task (to check for learning effects during the task) as fixed effects, and random intercepts for participants and items. This model showed an effect of the quantifier ( $X^2(2) = 11.9, p < .01$ ), the probability of an error being higher after *none* than after *some* ( $\beta = -2.21, SE = 0.6, z = -4.0, p < .001$ ), while the other contrasts were not significant ( $ps > .05$ ). The other fixed effects, including age, were not significant (all  $ps > .05$ ).

Here we turn to semantic vs pragmatic interpretations of the critical underinformative item, i.e., when *some* is used in a situation in which all the plates are initially filled. In this critical condition, pragmatic readings represent 79 % of responses overall, while semantic readings represent 20 %. The rate of pragmatic responses to critical items for each participant reveals a bimodal distribution (see Figure 4-3, left), with participants being disproportionately pragmatic. Results per class showed that the distribution was roughly evenly bimodal in 1<sup>st</sup> grade, with very few mixed profiles from 2<sup>nd</sup> grade on (see Figure 1 in [supplementary materials](#)). To investigate developmental progressions



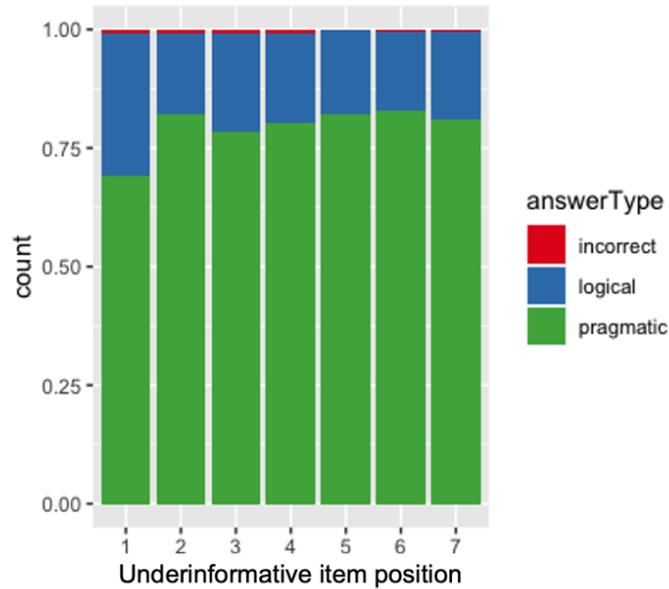
**Figure 4-3:** Distribution of the rates of pragmatic interpretations to critical items of each child (left) and their evolution with age (right: children’s individual pragmatic interpretation rates as a function of age, described with a LOESS regression line (plain yellow, SE in grey) and a linear regression line (dashed black))

of the underinformative item, we fitted a generalized logistic model that included all responses to the critical items (1708 data points), with the pragmatic response considered the dependent variable and age, SES and position of the item in the task as fixed effects. We added random intercepts for participants and items. Visual inspection of the data suggested that adding second order terms for SES and item position might improve the model, which was confirmed by the likelihood ratio tests (respectively,  $X^2(1) = 6.8$ ,  $p < .01$  and  $X^2(1) = 6.4$ ,  $p < .05$ ). Adding a second order term for age did not improve it ( $X^2(1) = 0.1$ ,  $p = .71$ ), indicating a linear progression. This final model confirmed that pragmatic interpretations are more likely to arise with age ( $\beta = 1.19$ ,  $SE = 0.31$ ,  $X^2(1) = 15.7$ ,  $p < .001$ ), the average rate of pragmatic interpretations increasing from around 65% among 6-year-olds to close to 100% among 11-year-olds (see Figure 4-3, right).

The model also revealed that SES had a significant positive non-linear effect on the probability of responding pragmatically ( $X^2(2) = 14.2$ ,  $p < .001$ ), with a large increase between 2 and 1 standard deviations below the mean and no

improvement after. Similarly, the position of the item in the task had a positive and non-linear effect ( $X^2(2) = 13.7$ ,  $p < .001$ ); that is, children were more likely to interpret *some* pragmatically at the end of the task than at the beginning, with most of this progression in the first half of the task. Data visualization (Figure 4-4) reveals that much of this effect seems explained by the difference between the first and the subsequent items.

In an exploratory perspective, we wanted to describe children’s tendency to operate changes on the scenes, independently from the instructions. To do so, we isolated control items for which changes were possible but not necessary, that is, items using the quantifier ‘some’ with a partial initial context (e.g. when the instruction was “*I want some of the plates to have a strawberry*” while 2 plates initially did: even if this was not necessary, children still had the possibility to add 1 or 2 fruits without changing the answer’s categorization). We excluded errors. Then we fitted a generalized model explaining the probability to make a change, with age as fixed effect and random intercepts for participants and items. This revealed that



**Figure 4-4:** Group responses to underinformative items for each item position; note that the first and the sixth item are identical

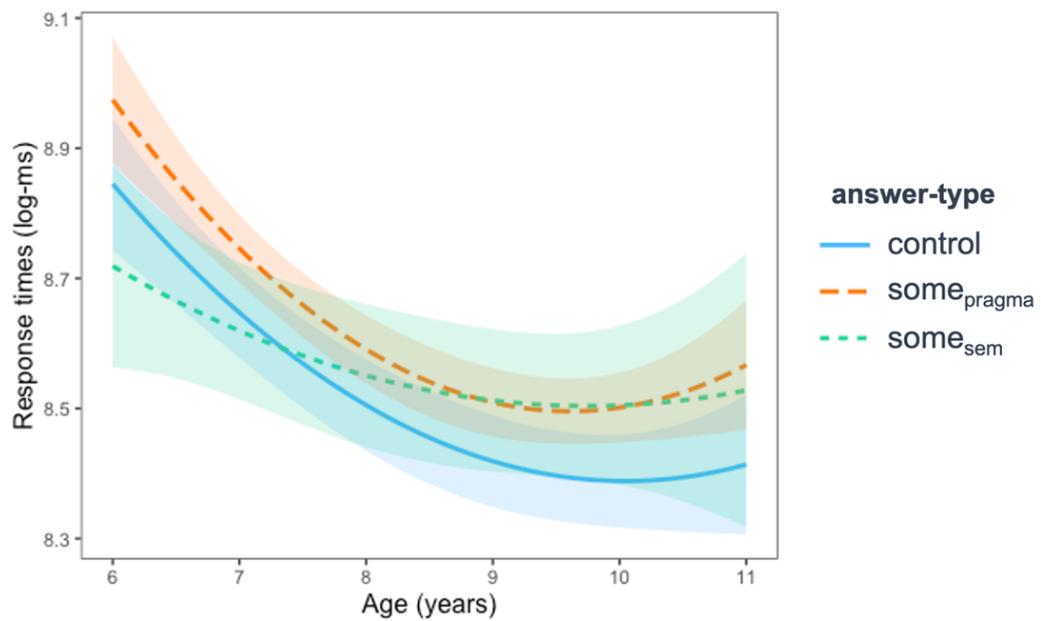
the probability to operate unnecessary changes was important but reliably decreased with age ( $\beta = -0.51$ ,  $SE = 0.12$ ,  $z = -4.1$ ,  $p < .001$ ), falling from approximately 70% at 6 years old to 30% at 11. Adding item position to this model did not improve its fit ( $X^2(1) = 0.009$ ,  $p = .93$ ).

### Response times (RT)

For response times analysis, we removed incorrect responses (1.9%) and repeated items, i.e., items in which the “repeat” button was used, allowing atypical strategies or unusual treatment of the target sentence (1.7% of the data). Then, we removed outliers for which log-transformed RT deviated more than 3 SD from the mean for each value of *moves* (1.5%). The final data set included 3475 data points.

To test our hypothesis that pragmatic responses are effort demanding, we fitted a model that accounts for log-transformed response times with the variables susceptible to influence response times as fixed effects. This includes *answer-type* (which includes control vs semantic readings of underinformative *some* vs pragmatic readings of underinformative

*some*), *age* (since children are likely to get faster with age), *item position*, because we expect them to get faster as they progress in the task, and the controls variables *hasMoved* and *moves*. We added random intercepts for participants and items. Visual inspection of the data showed that response times appear to have a non-linear relation with age; likelihood ratio test confirmed that adding a 2<sup>nd</sup> order term for age much improves the model’s fit ( $X^2(1) = 15.3$ ,  $p < 10^{-4}$ ). We did not retain a 3<sup>rd</sup> order term for age in the final model because adding it did not significantly improve the fit ( $X^2(1) = 0.002$ ,  $p = .96$ ). Then, we tested whether including the interaction of answer-type with item position improved the model (which was not the case,  $X^2(2) = 1.8$ ,  $p = .40$ ), as well as the interaction of answer-type with age, which was significant ( $X^2(4) = 13.7$ ,  $p < .01$ ). We thus kept that last interaction term in the final model, in which item position had a significant effect ( $X^2(1) = 20.1$ ,  $p < 10^{-5}$ ), as well as the two control variables, *hasMoved* ( $X^2(1) = 12.7$ ,  $p < .001$ ) and *moves* ( $X^2(1) = 12.7$ ,  $p < .001$ ). We examined the age by answer-type interaction by plotting the model’s predictions in Figure 4-5 and by computing



**Figure 4-5:** Predicted response times for each answer-type at different ages (partial effects plot, with 95% confidence intervals in shades)

post-hoc pairwise comparisons between the three answer-types at the exact age values of 6, 7, 8, 9, 10 and 11 years old (see all comparisons in [supplementary materials](#))<sup>19</sup>.

We first note that there was an overall age effect with a large acceleration of response times between 6 and 9 years old, and a relative tray after 9 years old (Figure 4-5). Then, pragmatic responses prove to be reliably slower than responses to control items (all  $p$ s < .05, except at 8 years old where  $p = .06$ ). Semantic responses showed a different pattern with age, being significantly faster than pragmatic responses at 6 and 7 years old (all  $p$ s < .05), but not after. At 9 and 10 years old, semantic responses tended to be slower than control responses, but this failed to reach significance ( $p < .10$ ). At 9 and 10, estimated differences were similar to the difference between pragmatic responses and controls, but with larger standard errors

(semantic responses were indeed much less frequent).

## Discussion

This action-based paradigm offers precious insight about children's interpretation of scalar terms. This is an important advance compared to the many prior studies that mostly used judgement tasks. The response time data provide hints about the processes that power these inferences, putting us in a position to address the debate between the *emerging pragmatic capacity* and the *pragmatic tolerance* accounts of the developmental effect. Before discussing our results, we summarize here the three main findings of our work. First, children's very low error rates across ages confirms that they are not struggling with the semantic meaning of the material used, nor with the

<sup>19</sup> This is not to say we used age as *groups*, but we kept age as a continuous predictor, following

(Royston et al., 2006) who argue that continuous predictors should not be dichotomized.

task design and implementation in general. Second, the rate of pragmatic responses to the underinformative items appeared higher than in most developmental studies, with an overall rate of 79% of pragmatic interpretations, but importantly, this rate progressively increased with age. Third, pragmatic readings of *some* were associated with longer response times compared to control responses. In younger children (6-7 years old), pragmatic responses were also slower than semantic responses, provided as fast as control responses. This difference seemed to fade away with age, and in older children semantic responses tended to take more time than controls, although this remained only marginally significant.

Although specific in some ways, our results fit well with the existing evidence on scalar implicature development. As Miller et al. (2005) and Pouscoulous et al. (2007), whose task ours is adapted from, we obtained higher rates of pragmatic responses than what can be observed with classical judgment task (Barner et al., 2011; Chierchia et al., 2004; Guasti et al., 2005; Noveck, 2001; Papafragou & Musolino, 2003). Pouscoulous et al. (2007) argued that using a response format that does not tap into metalinguistic skills might facilitate inference derivation; additional and effortful processes are indeed likely to be involved in sentence evaluation compared to direct interpretation as in our task. We also observed that our paradigm encourages children to make changes in the scene in general, even though they are not necessary. This general effect is likely to mechanically heighten the rates of responses categorized as pragmatic (where changes must be made) over semantic responses (where no change can be made). This might additionally explain why in this action-based task, pragmatic responses are more frequent than in judgment tasks.

Yet, we even observed higher rates of pragmatic responses than Pouscoulous et al.

(2007)'s original study. They obtained, in way smaller groups, 58% of pragmatic answers in 9-10 years old children and 79% in adults, while in our data the rate of pragmatic answers evolved from around 65% at 6 years old to a ceiling at 90% at 11 years old. One difference between our paradigms might contribute to this difference: they only presented one underinformative utterance to each child, whereas we used 7. Interestingly, adding item position as fixed effect to the model explaining inference derivation improved its fit, and visual inspection of the data revealed that this effect was mostly due to the very first underinformative item which received fewer pragmatic responses than other items, at all ages. If we only considered the first underinformative item, our rates would be closer to those expected from Pouscoulous et al. (2007). Note, however, that the task we used is part the *TIPi*, a battery designed for clinical use, which requires a unique item list: the item position effect observed thus can't be distinguished from an item effect, even though there is no *a priori* reason why this specific item should be harder than any other, especially since the first and the sixth underinformative items were identical. Even though we gave no feedback during the task and that no effect of item position was observed on error's probability, this result could be interpreted as a sort of training effect. The effort demanded by the scalar inference seems higher at the first presentation than the others, either because the inference process is then pre-activated for the following items or because the linguistic alternative to *some* (namely, *all*) has been presented multiple times, becoming more available as the task goes on. Previous studies already showed that making the alternative more accessible helped children derive scalar inferences (e.g., Bagassi et al., 2009; Skordos & Papafragou, 2016). This result is also very consistent with Ronderos & Noveck (2023)

study, who also observed in adults, using response times, that scalar implicature derivation was getting easier as the task goes on. However, in our results, we failed to replicate such effects on response times: while children got faster as they move on, they did so independently of response types. It is likely that the high variability of response times in our paradigm would prevent us from detecting such subtle effect.

Crucially, pragmatic answers were clearly more likely to arise with age, as expected. This is all the more significant since, as argued earlier, the task design itself favors pragmatic responses especially in young children, who are those who appear to be the less likely to produce them. We thus replicated a classical finding in a large sample, adding fine-grained information about the progression of pragmatic development in school-aged children. While other studies usually compare different age groups, our sample size allows the computation of a continuous and reliable relation between age and pragmatic responses. That progression appeared linear, with no particular leap at any age. The bimodality of the individual rates distribution is also consistent with previous studies such as Guasti et al. (2005). In our sample, this bimodality could be observed as early as in 1<sup>st</sup> graders. The developmental effect can thus be seen as an increase of the likelihood of a child to be a pragmatic responder, rather than an increase of the number of pragmatic answers that one child provides with age. Again, this contrasts with the rarity of semantic errors and its independence from age, which indicates that the semantic meaning of scalar term is not challenging for school age children.

The developmental trend observed with our action-based paradigm, which *directly* reveals how children interpret scalar terms, is thus the same as what could be expected from judgment tasks results. This indicated

that the latter reflect (at least partly) inference derivation, and not only a response strategy, despite the reasonable methodological criticism made by Katsos & Bishop (2011). On the contrary, our results contrast with the results they obtained with a ternary judgment task, where 5-6 years-old children showed adult-like performances. This paradigm deploys great efforts to focus the participants' attention on the linguistic form and its felicity by (1) explicitly instructing participants to do so, (2) explaining that the speaker does not perfectly master the language, and (3) providing a response format that encourages metalinguistic reasoning rather than a mere direct interpretation. Thus, we interpret Katsos and Bishop's task as a measure of how sensitive to informativity a participant *can be when they are explicitly asked* to, rather than an index of the *spontaneous* use of informativity to infer a speaker's meaning. In contrast, we argue that asking a participant whether *some elephants have trunks* or not might reveal a more spontaneous interpretation of the sentence (i.e., it is more likely to reflect the actual inference derivation), even though the methodological criticism holds. This would explain why our paradigm and sentence judgment task provide consistent results.

Additionally, our data provide supplementary insights on implicature processing with the response times of children. Before turning to the theoretical debate these data can address, we first rule out the impressive acceleration of children's responses with age, response times dropping from about 8000 ms at 6 years old to almost a half, at approximately 4500 ms at 9 years old and after. Such a change contrasts with performances, where no major turning point could be observed in any case. Even if young children were much slower than older ones, they did not commit more errors. Therefore, we attribute this

general acceleration to the development of children's ability to efficiently manipulate the tablet-based material and to focus generally on the task. Additionally to this strong and reliable effect, response times varied across our condition of interest, and this effect was modulated by age (see Figure 4-5). In younger children, pragmatic responses took longer than semantic readings, which were provided as fast as control responses. However, the difference between pragmatic and semantic responses decreased with age, while semantic and control responses progressively tended to dissociate, yet failing to reach significance. For example, at 7 years old (all other effects being accounted for), pragmatic responses were estimated to be provided with a latency of 5 653 ms, contra 5 115 ms for control responses and 4 964 ms for semantic responses. At 10 years old, pragmatic responses arose after 4 402 ms and semantic responses after 4 447 ms, contra 3 944 ms for control responses.

In accordance with our hypothesis, the pattern observed in younger children is thus comparable to the data reported by Bott & Noveck (2004) with adult participants (recently replicated in Ronderos & Noveck [2023]), even though the paradigm was not exactly the same. However, this was not the case for older children. We consider two different explanations for that dissociation. First, semantic responses were much less frequent in older children, as described earlier; this led to wider standard errors, as visible in Figure 4-5, which hindered the power of statistical analysis. This is especially true with our paradigm whose advantage of being naturalistic and unconstrained come with a cost in terms of RT precision, inducing a lot of noise and requiring a lot of participants and data points. It is thus possible that our model might have failed to capture meaningful differences between semantic responses and the other responses in this part of our

sample. The alternative (and not necessarily exclusive) interpretation is that the processing mechanisms that underlie responses to our task could be evolving with age, explaining the reversing trend that emerged. To discuss this possibility, we need to turn to the theoretical debate that we wanted to address, namely the opposition of the *emerging pragmatic capacity* and *pragmatic tolerance* accounts.

It is worth to focus first on the pattern observed in younger children. We predicted that if participants were spontaneously sensitive to informativeness but tolerant with violations, underinformative items should always be more complex to process, no matter what response is provided (semantic or pragmatic). Indeed, as stated by Katsos and Bishop (2011), an underinformative utterance is ambiguous, being both true (regarding semantic content) and false (regarding felicity); if perceived, this ambiguity must be solved to reconstruct the communicative intention undermining the utterance. With our task, we observed that 6-7 years old children were slower when they interpreted *some* pragmatically, than when they gave semantic interpretations, which were in turn provided as fast as control items. A *pragmatic emerging capacity* but not *pragmatic tolerance*, would predict these observations. When semantic responders are presented with underinformative *some*, they process it as fast as they do with other quantifiers, as if they did not grasp the ambiguity that pragmatic responders seem to have to deal with. One way to interpret together our data and Katsos and Bishop's (2011) would be to say that young children are *able* to process informativeness when they are asked to and pushed to it, as by the ternary judgment task, but that they struggle to spontaneously use informativeness violation to derive the speaker's meaning when presented with underinformative sentences in naturalistic situations.

This confirmed developmental evidence goes together with a variety of methodologies used with adults to point the cost of scalar inferences, at the cognitive level using eye tracking (Huang & Snedeker, 2009, 2018), reading times (Breheny et al., 2006) or cognitive pressions paradigms (De Neys & Schaeken, 2007) as well as at the cerebral level with techniques like EEG (Zhao et al., 2021) or fMRI (Shetreet et al., 2014; Zhan et al., 2017). The effortful nature of scalar inferences would explain why children struggle to spontaneously execute them, while they can detect underinformativity when they are pushed to. The origin of that effort can be linked to the general cognitive resources needed to generate alternatives to the uttered message and negate them to update the speaker's mental states which are simultaneously tracked. This fits well with the general framework of relevance theory (Sperber & Wilson, 1996) which posits that communicative effects and pragmatic inferences come with a cost, in terms of cognitive resources. More specifically, some cognitive functions are known to be still maturing in this age range and have been posited to be involved in pragmatic inferences, especially theory of mind (Fairchild & Papafragou, 2021; Ronderos & Noveck, 2023) and executive functions (Bambini et al., 2021). Be it the general cognitive resources or more late developing functions, it would explain that in our results, children become continuously more pragmatic in school age, with no particular threshold or turning point. This is consistent with other results showing that lowering the task cognitive demand in general improves pragmatic performance, which can be done by using more frequent lexical items (Pouscoulous et al., 2007, Experiment 3), by training children before testing (Guasti et al., 2005; Papafragou & Musolino, 2003), or by using an action-based paradigm like ours.

This being said, these our interpretations must be nuanced by the age by response-type interaction observed in our data. Even though semantic responses are never significantly longer than control responses, a trend is observed toward that direction in older children. It is thus possible that, as they grow, children become more spontaneously sensitive to informativity, even when they (more rarely) keep answering in a semantic way. The nature of semantic responses might thus be evolving: while they would reflect a pragmatic immaturity in younger children, they would be better explained by a response strategy in older children who could be aware of these utterances' ambiguity, that is, by pragmatic tolerance. Further studies should be run to rule out this possibility, with paradigms that prompt more semantic responses than ours, providing more data point and strengthening statistical analysis.

One possible way to go further and disentangle the nature of the effort engaged though pragmatic inference would be to turn to atypical populations, such as autism spectrum disorders. In this domain, most studies expected but failed to observe clear differences in the treatment of scalars between autistic adults, teenagers or children and typically developing pairs, especially when groups were matched for IQ (Chevallier et al., 2010; Hochstein et al., 2017; Pijnacker et al., 2009; Su & Su, 2015). But importantly, all these studies used sentence judgment tasks which, as discussed earlier, do not necessarily reflect how participants spontaneously interpret underinformative sentences. Even though we argued that *in typical development* judgment tasks do approximate participants' actual interpretations of scalar term, we still acknowledge Katsos' (2014) methodological point. We believe that this approximation might be different in atypical populations and thus join Katsos' call for a variety of tasks, such as the

response times action-based paradigm we used.

## Conclusion

Much work has been done to study how scalar implicatures develop in typical childhood and how they were processed. Different studies have suggested that children were *immature* in interpreting such utterances until late, yet mostly with indirect measures of inference derivation that are judgement tasks. In this study, we used an innovative action-based paradigm which provided a direct measure of children's interpretation of scalars and the associated response times. Our results are mostly in line with the *emerging pragmatic capacity* claim. Together with the data supporting the *pragmatic tolerance* account, however, our study suggests that this immaturity can't be explained by a general insensitivity to informativeness but to a lack of spontaneous use of informativity to derive the speaker's meaning (modulated by the experimental paradigm used), at least in 6-7 years old children. The reasons for this immaturity, most likely linked to the cognitive complexity of this operation, are still debated and need further investigations. Moreover, our results suggested that a change in the processes underlying responses could happen with age. It is possible that *pragmatic tolerance* explains better the semantic responses of older children (of 9-11 years old), but further evidence should be provided to support that claim. Future research could benefit from the collection of data in various populations and the use of a diversity of tasks, including the innovative paradigm we proposed.

## References

Bagassi, M., D'Addario, M., Macchi, L., & Sala, V. (2009). Children's acceptance of

underinformative sentences: The case of *some* as a determiner. *Thinking & Reasoning*, *15*(2), 211–235.

<https://doi.org/10.1080/13546780902864306>

Bambini, V., Van Looy, L., Demiddele, K., & Schaeken, W. (2021). What is the contribution of executive functions to communicative-pragmatic skills? Insights from aging and different types of pragmatic inference. *Cognitive Processing*, *22*(3), 435–452. <https://doi.org/10.1007/s10339-021-01021-w>

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93. <https://doi.org/10.1016/j.cognition.2010.10.010>

Bott, L., & Noveck, I. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457. <https://doi.org/10.1016/j.jml.2004.05.006>

Boyce, W., Torsheim, T., Currie, C., & Zambon, A. (2006). The Family Affluence Scale as a Measure of National Wealth: Validation of an Adolescent Self-Report Measure. *Social Indicators Research*, *78*(3), 473–487. <https://doi.org/10.1007/s11205-005-1607-6>

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463. <https://doi.org/10.1016/j.cognition.2005.07.003>

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar Inferences in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *40*(9), 1104–1117. <https://doi.org/10.1007/s10803-010-0960-8>

Chierchia, G., Guasti, M. T., Gualmini, A., Meroni, L., Crain, S., & Foppolo, F. (2004). Semantic and pragmatic competence in children's and adults' comprehension of or. *Experimental Pragmatics*, 283–300.

Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) Family Affluence Scale. *Social Science & Medicine*,

- 66(6), 1429–1436. <https://doi.org/10.1016/j.socscimed.2007.11.024>
- De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, *54*(2), 128–133.
- Fairchild, S., & Papafragou, A. (2021). The Role of Executive Function and Theory of Mind in Pragmatic Computations. *Cognitive Science*, *45*(2). <https://doi.org/10.1111/cogs.12938>
- Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.
- Guasti, T. M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667–696. <https://doi.org/10.1080/01690960444000250>
- 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>
- Horn, L. R. (1989). *A natural history of negation* (University of Chicago Press).
- Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723–1739. <https://doi.org/10.1037/a0016704>
- Huang, Y. T., & Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cognitive Psychology*, *102*, 105–126. <https://doi.org/10.1016/j.cogpsych.2018.01.004>
- Katsos, N. (2014). Scalar implicature. *Pragmatic Development in First Language Acquisition*, *10*, 183–197.
- Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81. <https://doi.org/10.1016/j.cognition.2011.02.015>
- Katsos, N., Cummins, C., Ezeizabarrena, M.-J., Gavarró, A., Kuvač Kraljević, J., Hrzica, G., Grohmann, K. K., Skordi, A., Jensen de López, K., Sundahl, L., van Hout, A., Hollebrandse, B., Overweg, J., Faber, M., van Koert, M., Smith, N., Vija, M., Zupping, S., Kunnari, S., ... Noveck, I. (2016). Cross-linguistic patterns in the acquisition of quantifiers. *Proceedings of the National Academy of Sciences*, *113*(33), 9244–9249. <https://doi.org/10.1073/pnas.1601341113>
- Mazzaggio, G., Foppolo, F., Job, R., & Surian, L. (2021). Ad-hoc and scalar implicatures in children with autism spectrum disorder. *Journal of Communication Disorders*, *90*, 106089. <https://doi.org/10.1016/j.jcomdis.2021.106089>
- Miller, K., Schmitt, C., Chang, H.-H., & Munn, A. (2005). Young children understand some implicatures. *Proceedings of the 29th Annual Boston University Conference on Language Development*, 389–400.
- Noveck, I. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188. [https://doi.org/10.1016/S0010-0277\(00\)00114-1](https://doi.org/10.1016/S0010-0277(00)00114-1)
- Noveck, I. (2018). *Experimental Pragmatics: The Making of a Cognitive Science*. Cambridge University Press. <https://doi.org/10.1017/9781316027073>
- Noveck, I., & Sperber, D. (Eds.). (2004). *Experimental Pragmatics*. Palgrave Macmillan UK. <https://doi.org/10.1057/9780230524125>
- Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253–282. [https://doi.org/10.1016/S0010-0277\(02\)00179-8](https://doi.org/10.1016/S0010-0277(02)00179-8)
- Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J.-P., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *39*(4), 607–618. <https://doi.org/10.1007/s10803-008-0661-8>
- Pouscoulous, N., Noveck, I., Politzer, G., & Bastide, A. (2007). A Developmental Investigation of Processing Costs in Implicature Production. *Language Acquisition*, *14*(4), 347–

375.

<https://doi.org/10.1080/10489220701600457>

Ronderos, C. R., & Noveck, I. (2023). Slowdowns in scalar implicature processing: Isolating the intention-reading costs in the Bott & Noveck task. *Cognition*, *238*, 105480.

Shetreet, E., Chierchia, G., & Gaab, N. (2014). When some is not every: Dissociating scalar implicature generation and mismatch: fMRI of Scalar Implicature Mismatch. *Human Brain Mapping*, *35*(4), 1503–1514. <https://doi.org/10.1002/hbm.22269>

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*, 6–18. <https://doi.org/10.1016/j.cognition.2016.04.006>

Sperber, D., & Wilson, D. (1996). *Relevance: Communication and Cognition* (2nd ed.). Wiley-Blackwell.

Stephens, G., & Matthews, D. (2014). The communicative infant from 0-18 months: The social-cognitive foundations of pragmatic development. In *Pragmatic development in first language acquisition* (pp. 13–35). John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.10.02ste>

Su, Y., & Su, L.-Y. (2015). Interpretation of Logical Words in Mandarin-Speaking Children with Autism Spectrum Disorders: Uncovering Knowledge of Semantics and Pragmatics. *Journal of Autism and Developmental Disorders*, *45*(7), 1938–1950. <https://doi.org/10.1007/s10803-014-2350-0>

Zhan, J., Jiang, X., Politzer-Ahles, S., & Zhou, X. (2017). Neural correlates of fine-grained meaning distinctions: An fMRI investigation of scalar quantifiers: Neural Correlates of Fine-Grained Meaning Distinctions. *Human Brain Mapping*, *38*(8), 3848–3864. <https://doi.org/10.1002/hbm.23633>

Zhao, M., Liu, X., Dai, X., Dong, S., & Han, Z. (2021). Scalar implicature is not a default process: An ERP study of the scalar implicature processing under the effect of focus factor. *Brain Research*, 147499. <https://doi.org/10.1016/j.brainres.2021.147499>

## Point d'étape à l'issue de l'étude 1

Comme discuté au [chapitre 2](#), les implicatures scalaires sont classiquement considérées comme difficiles d'accès pour les enfants, en raison d'une immaturité des processus pragmatiques. Toutefois, cette conclusion a principalement été tirée des résultats de tâches de jugements binaires, qui ne reflètent pas nécessairement la compréhension de ces implicatures. De plus, ces résultats pourraient refléter une tolérance pragmatique vis-à-vis d'énoncés sous-informatifs, plutôt que d'une immaturité propre des habiletés pragmatiques. Dans cette étude, nous avons donc mis en place un paradigme expérimental sur tablettes permettant d'évaluer *directement* comment des enfants interprètent un quantifieur scalaire ('*certain*'). Cette tâche, incluse dans le TIPi, un nouvel outil d'évaluation des inférences pragmatiques à des fins cliniques, permet une réponse par l'action, et l'enregistrement des temps de réponse des participants. L'inclusion d'un large échantillon de 249 enfants à développement typique de 6 à 11 ans permet en outre d'obtenir des indications sur la trajectoire développementale précise de cette compétence, par rapport à la comparaison de groupes d'âges arbitrairement délimités. Les résultats de cette étude confirment l'effet développemental classiquement observé, révélant une augmentation progressive et linéaire de la probabilité d'interpréter pragmatiquement un terme scalaire avec l'âge. De plus, les temps de réponses révèlent que ces interprétations pragmatiques sont plus coûteuses que des interprétations sémantiques, qui sont réalisées aussi vites que dans le cas des situations contrôles, au moins chez les enfants de 6-7 ans. Ces interprétations pragmatiques reflètent ainsi probablement une interprétation plus riche et mature que ces autres situations, suggérant que les habiletés pragmatiques qui les sous-tendent sont encore en développement à l'âge scolaire. À l'inverse, l'hypothèse selon laquelle les enfants seraient précocement sensibles à l'informativité d'un énoncé mais seraient plus tolérants vis-à-vis d'infractions à ce niveau paraît moins compatible avec nos données. Dans l'ensemble, ces résultats et les études précédentes suggèrent que les enfants sont en mesure d'évaluer l'informativité d'un énoncé lorsqu'on leur demande explicitement, mais qu'ils peinent à utiliser spontanément cette information pour inférer les intentions de communication d'un locuteur. Cette étude précise donc la nature des développements pragmatiques qui sont à l'œuvre à l'âge scolaire, pour ce qui est des implicatures scalaires. Dans [l'étude suivante](#), nous mettrons en place une démarche similaires visant à préciser ces développements dans le domaine des métaphores, en répondant aux limites de plusieurs études antérieures, et en élargissant nos investigations aux interactions entre ces développements et ceux du langage et de la théorie de l'esprit. Nous pourrions à cette fin exploiter une tâche informatisée et non-verbale de théorie de l'esprit que nous avons adaptée d'une tâche classique de Langdon & Coltheart (1999) et validée pour une utilisation en clinique dans une étude annexe ([annexe B](#)). En outre, les résultats de cette première étude sur le développement typique pourront être directement comparés aux performances d'enfants autistes dans les données cliniques préliminaires présentées dans le [chapitre 6](#).

## Chapitre 5 • Développement typique des métaphores (étude 2)

### Article *en révision* pour le *Journal of Experimental Psychology: Learning, Memory, and Cognition*: How do theory of mind and language skills impact metaphoric reference comprehension during children's school-age years

Nicolas PETIT, Valentina BAMBINI, Luca BISCHETTI, Jérôme PRADO, Ira NOVECK

---

#### Abstract

Metaphor comprehension development, a long-standing topic in the cognitive science, is the source of multiple strands of research. These provided evidence both for the presence of early precursors in young children and for a progressive and relatively late mastery of metaphor, which this work focuses on. Moreover, research in the field of atypical development has put to the fore two important cognitive bases for metaphor comprehension, that are theory of mind (TOM) and structural language, but these have remained relatively understudied in typical development, and not yet in a joint perspective. To address this issue, this work introduces a new tablet-based, orally presented metaphoric reference task, which allows for response times collection. In Experiment 1, we confirmed the viability and developmental sensitivity of this task in typically developing 7- and 10-year-old children (N=89). In Experiment 2, we used a version of this task with 248 typically developing 6- to 11-year-olds, together with language and non-verbal TOM tasks. Results showed a non-linear developmental trend in metaphor comprehension, with a focused improvement between 7 and 9 years of age. Mixed-effect models explaining children's accuracy and response times revealed a critical developmental shift with respect to the cognitive predictors of metaphor comprehension. The data showed that TOM is a reliable predictor of successful performance among younger children but fades away, at which point language becomes a predictor of success on the task among older children. This study underlines the importance of considering developmental perspectives when referring to the cognitive bases of metaphor skills.

---

Research on the development of metaphor comprehension can be reduced to three overlapping, but largely-independent, strands of investigations. Historically speaking, the first of the current strands concerns the development of metaphor comprehension among school-aged

children. This line of work began with Piaget (1964) who claimed – based on findings with out-of-context proverbs (which are often but not always metaphorical) – that children need to reach adolescence before they could understand and produce metaphoric utterances.

Piaget's account would eventually spur many researchers to report observations – often from their own children – contradicting his claim (see e.g., Gentner, 1988) and it prompted many others to carry out developmental studies showing that young children are indeed much more competent with metaphor than he proposed (e.g., Vosniadou, 1987). As research studies accrued, the findings have eventually revealed support for two complementary claims: one is that metaphor comprehension indeed continues to improve with age, to the point that children become adult-like progressively. However, research also shows that there is some rudimentary ability among children who are as young as 3 years of age to use and understand metaphor (Pouscoulous, 2011; Pouscoulous & Tomasello, 2020). This accounts for the first two of the strands.

Along the way, the *nature* of metaphor itself became central to investigations. That is, researchers took an interest in children's metaphor performance in order to a) investigate the extent to which the properties of metaphor could account for children's difficulties and to; b) better understand the underlying abilities that facilitate metaphor comprehension. This is what ultimately led to the third strand. To make this clear, consider two examples that underline the importance of psychological factors in metaphor comprehension. The first comes from Winner et al. (1976) who showed how psychological metaphors, e.g., *The prison guard had become a hard rock*, are more difficult for children to understand than physical metaphors, e.g., *hair is spaghetti* (for more recent results in this vein, see Lecce et al., 2019). The second, and arguably more impactful, example in this line comes from developmental research showing that figurative language (such as metaphor) could be impaired in autistic compared to neurotypical participants (Happé, 1993; for reviews see

Kalandadze et al., 2018 or Morsanyi et al., 2020). The proposal motivating Happé's work was that theory of mind (TOM, the ability to understand an interlocutor's intended meaning) is a critical factor for metaphor comprehension. Once researchers considered a role for theory of mind in metaphor comprehension, it became only natural to test this claim against alternative factors that could account for developmental changes, such as structural language abilities (Norbury, 2005).

By viewing the metaphor development literature in this three-pronged way, it becomes clearer why it is populated with such a wide variety of studies. Moreover, by appreciating that there are multiple strands, one can better appreciate how the literature has evolved. Given the breadth of investigations on metaphor development, it would obviously be difficult for any single study to address all three strands. It is nonetheless worthwhile to attempt to interlace, at the very least, two of them. Here, we introduce a task that investigates the development of metaphoric (versus synonymic) referential comprehension during school age, which is tested through participants' choice of picture in the wake of a referential expression, and we go on to consider the relative influence of two underlying factors that are often assumed to be critical to metaphor: theory of mind and structural language comprehension.

We break down the remainder of this Introduction into three parts. First, we summarize studies that investigate school-aged children with tasks that endeavor to capture developmental progress with metaphor. Second, we briefly review findings that aim to uncover the substrates of metaphor comprehension. This will lead to a discussion concerning the way theory of mind as well as language abilities affect participants' growing metaphor comprehension. Finally, we will introduce our two Experiments, which essentially

investigate children's accuracy in identifying a metaphoric reference while listening to orally-presented vignettes via a tablet.

### **The development of metaphoric abilities among school-aged children**

A large number of studies now assume that school-aged children do in fact progressively gain competence with metaphor with age. For the purposes of providing background on our study, we describe two streams of research that have been particularly helpful in providing insights into metaphor development. The first relies on a paradigm that asks children to *detect* and metalinguistically *explain* metaphoric relations among triads. Kogan et al. (1980) introduced what they called the *Metaphor Triad Task* (the MTT), in which young participants are presented with three items that allow for up to three pairing possibilities, one of which is metaphoric in character while the remaining two reflect other sorts of relationships. In this task, participants are requested, by an experimenter, to select the best possible pairing and to explain it. They are also asked to continue (and are thus slightly prodded) so that they ultimately always consider the metaphoric one. Consider the triad *fish*, *winding river*, and *snake*. Here, the latter two would be expected to be the source of metaphoric pair. This technique captured children's growing metaphoric abilities reliably whether the items in the triad were presented pictorially or with words (Kogan & Chadrow, 1986).

To appreciate this technique in a more recent study, consider work from Willinger's group (Deckert et al., 2019; Willinger et al., 2019). Inspired by the prior work, Willinger et al. (2019) tested 164 seven-, nine- and eleven-year-olds with a form of the MTT and recorded three different scores for each triad: whether the metaphorical relation was identified or not

(*identification*), whether it was correctly explained or not (*explanation*), and whether it was preferred over other associations, i.e., whether it was identified first (*preference*). Their findings revealed that the eleven-year-olds had higher metaphor comprehension quality and preference scores than the younger groups. Deckert et al. (2019) tested 298 typically developing children, between the ages of 7 and 11, and observed different turning points, depending on the measure they considered. A first turning point was identified at 8.2 years of age, after which children *identified* the metaphoric pairs and *explained* them better, and a second was observed at 10.2 years old, after which they increasingly *preferred* metaphoric pairs over non-metaphoric ones. This resonated with Willinger and colleagues' (2019) observations. The authors explain that these gaps might correspond to changes in the *nature* of the processing strategies of metaphors – and not to a continuous development – that they link to brain structure modifications which occur over the course of childhood. Willinger et al. (2019) also linked this change to the concurrent increase of the communicative experience, at an age at which relationships are known to become more complex.

A second stream of research that investigates school-aged children's increasing abilities with metaphor comes from studies that use vignettes that ultimately include metaphoric lines of text that refer back to a previously introduced element, a technique that was initially introduced with adults by Gibbs's (1990). For example, in one vignette, Gibbs (1990) used the word *creampuff* in the sentence *The creampuff didn't even show up* to refer metaphorically to a previously mentioned boxer. Noveck et al. (2001) employed this sort of referential paradigm to investigate the development of metaphor comprehension. In their first experiment,

Noveck et al. (2001) presented 8- to 12-year-old French-speaking children with written vignettes whose penultimate sentence included a reference to a previously mentioned element. Consider one of their stories about school pupils in a pool who are later referred to as either “toads” metaphorically or else as “children” synonymically. The authors then determined, through yes/no questions, whether participants understood the reference (e.g., “was it the pupils who were sent to the side of the pool?”). They observed that, while referential abilities improved with age in general, metaphoric references consistently prompted more errors than the synonymic ones, until around 12 years old, where the gap appears to close. In a second experiment, they implemented this paradigm in the form of a self-paced computer task, in order to capture reading times for the referential sentences. Their data revealed – for children as well as adults – that metaphoric references were associated with longer reading times than their synonymic controls; this indicates that there is a cost in making metaphoric references. Similar referential tasks have been used in other developmental studies (Seigneuric et al., 2016; Tonini et al., 2023) and led to similar studies with adults that confirmed a temporal cost associated with metaphor processing (Almor et al., 2007; Heredia & Cieslicka, 2015; though see Carston & Yan, 2022).

Van Herwegen et al. (2013) designed a similar reference task in which metaphors (as well as metonymies) were ultimately embedded in orally presented stories that were six or seven sentences in length and that were accompanied by drawings along the way. At the end of each story, a question was presented to children about the meaning of a metaphor used in the story. The participants were presented with three new drawings and answered by selecting

one: one picture represented the intended object of the metaphor (e.g., a car which was described as a *turtle*), one represented a literal interpretation of the metaphorical expression (a turtle), and the third was a distractor. The 3 to 17 years-old participants’ picture-choice results showed that metaphor comprehension increased significantly with age.

Taken together, these studies show that children’s competencies with metaphor grow over time. Whether it concerns justifying the pairing of words or the comprehension of stories, metaphoric competence appears to increase during children’s school age years. Another question then becomes what motivates participants’ progress?

### **On the substrates of metaphor comprehension among children**

Here, we turn to the variability of children’s performance across studies as a function of task properties (for a recent review, see Kalandadze et al., 2019). As we indicated above, one of the earliest proposals concerning the predictors of children’s metaphor comprehension was through the prism of autism, which is known to be associated with a more literal understanding of language. This was what led Happé (1993) to link autistic individuals’ difficulties in metaphor comprehension to their presumed impairment with respect to theory of mind. Indeed, in line with Relevance Theory (Sperber & Wilson, 1996), accessing the meaning of a metaphor is achieved through inferences about the speaker’s intention.

One recent study that investigates the role of TOM in metaphor comprehension development comes from Lecce et al. (2019), who introduced a new Physical and Mental Metaphor (PMM) task while using metaphors of the form “X is Y.” Critical to the study was a comparison of *physical*

metaphors, which rely on features such as appearance or behavior (e.g., *dancers are butterflies*) to *mental* metaphors, which rely on psychological traits (e.g., *daddy is a volcano*). In their task, 217 9- to 12-year-old children were aurally presented with metaphors and asked to explain their meaning. Two different measures were coded from each response: whether the explanation was acceptable (*accuracy*) and whether the explanation relied on physical or mental features (*interpretation*). Results revealed that *accuracy* scores were high overall and, interestingly, that regarding *interpretation* scores, the nine-year-olds, the youngest group in the sample, performed at rates that were inferior to the other age-groups on mental, but not physical, metaphors. That is, while children provided plausible explanations overall, older children appear to rely on psychological features more than the youngest children, leading to a clear developmental effect between 9 and 11 years of age. Children from this study were also directly tested with the Strange Stories task (Happé, 1994), a now classic advanced TOM measure. Interestingly, they observed that this TOM measure predicted mental (but not physical) metaphor interpretation at 9 years of age, after controlling for vocabulary, SES and working memory, but they found no such associations for 10- and 11-year-olds. They thus introduced the idea that TOM's influence on metaphor interpretation could be short-lived as children grow (see also Tonini et al., 2023). Accordingly, in a short-term longitudinal study, the same team would go on to observe an association between metaphoric explanations and theory of mind growth among 9-year-olds (Del Sette et al., 2020), confirming the likely importance of TOM for children's metaphor comprehension at that age.

As we indicated above, Happé's seminal work led to an alternative proposal from

Norbury (2005), who argued that formal language skills were an important basis for metaphor comprehension. According to this view, semantic ability would be a better predictor of children's metaphor comprehension, since the semantic traits of the vehicle (e.g., the metaphoric meaning of *shark* in *my lawyer is a shark*) that need to be connected to those of the topic (*lawyer* in this example) are often not the most salient traits; this implies that sophisticated semantic representations are called for when discerning metaphors. Although work from these two seminal streams (the one centered on TOM and the other centered on semantic meaning or language) prompted some criticism, these abilities have emerged as the two most dominant candidates for the cognitive scaffolding of metaphor comprehension in the field of autism (for a review see Kalandadze et al., 2018) as well as for neurotypical development.

While considering the relative importance of TOM versus language abilities, Whyte & Nelson (2015) tested a group of 69 typically developing children between the ages of 5 and 12, on non-literal language (which included not only metaphor, but other pragmatic phenomena such as sarcasm and indirect requests). They observed that both theory of mind, measured with the Reading the Mind in the Eyes test (Baron-Cohen et al., 2001) and language, as measured with tests of vocabulary and syntax, are correlated with increasing comprehension. Based on Whyte & Nelson, one can say that the two abilities – TOM and linguistic abilities – have a role to play in children's growing abilities with non-literal language, but it is not clear whether the two work in tandem or separately. Thus far, no research, as far as we know, has aimed to disentangle the two in a single developmental study.

That said, researchers do occasionally try to separate out the influence of key abilities on metaphor development. For example, the previously-mentioned Seigneuric et al.

(2016) study used a written referential metaphor task in school-aged children to describe how “poor” comprehenders (defined as children with poor text comprehension but average or above average decoding abilities and vocabulary) had lower accuracy rates when identifying metaphoric referents, as opposed to literal referents. This led them to conclude that comprehension difficulties are due to something other than decoding and vocabulary. Similarly, in Deckert et al.'s (2019) metaphor triad task described earlier, the authors reported (page 347) that “for age groups seven and eight, [metaphor] comprehension was not significantly predicted by verbal intelligence and linguistic competence [...] whereas in age-groups nine and ten comprehension was significantly predicted by verbal intelligence.” In other words, verbal intelligence accounted for increases in metaphor comprehension among 9- and 10-year-olds more than among 7- and 8-year-olds. Data like these underscore the possibility that metaphor comprehension does not occur steadily and that specific competencies could have a passing influence on the development of metaphor comprehension, i.e., TOM and language skills could impose themselves on metaphor comprehension at specific points in a child’s cognitive life.

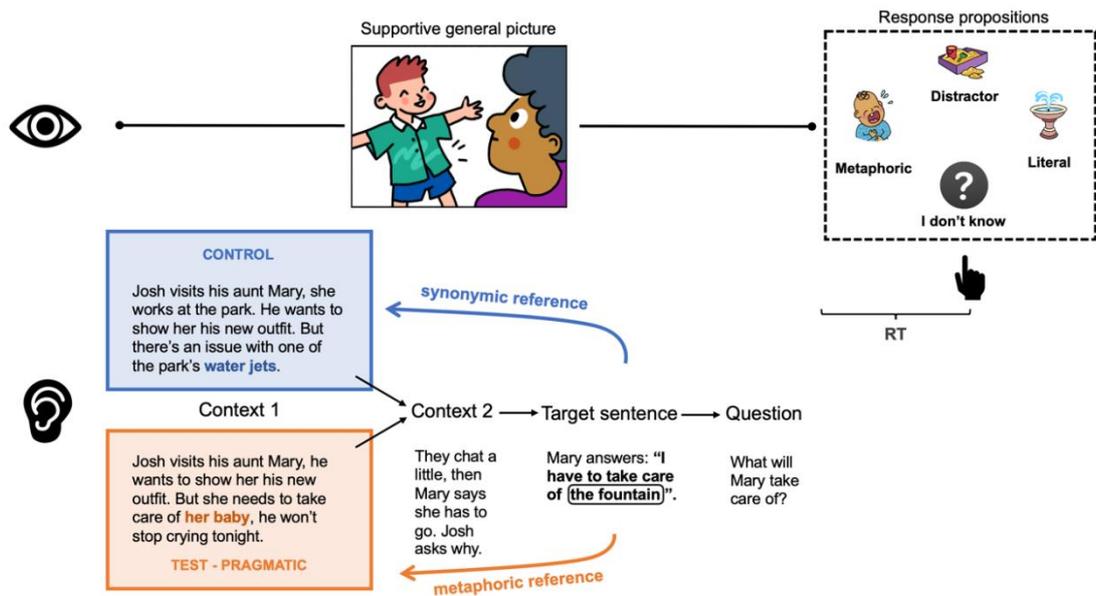
### **Goals of the current study**

The current work has two general goals. One is to confirm that a new variations of a self-paced metaphor-reference comprehension task, as developed here, provides developmental outcomes that are consistent with prior studies (e.g., Noveck et al., 2001; Seigneuric et al., 2016; Van Herwegen et al., 2013). This is important because the new variations to be used here amounts to a novel task. In each task item, participants listen to a vignette that introduce a narrative element (e.g., *a crying*

*baby*), that will be referred to in the target sentence, at the end of the vignette. Crucially, this reference can be metaphorical (say, ‘*the fountain*’ is used to refer to the crying baby), or else synonymic, as a function of the vignette opening context (‘*the fountain*’ could also be used to refer to some *water jets* that were introduced instead of a crying baby). Then, participants are asked a question to be answered through a picture selection, that can be timed and thus reveal how well and how fast they comprehend this reference. Importantly, in the metaphor condition, the intended target (a crying baby) is represented alongside a foil, which is a pictorial representation of the reference itself (a fountain, see **Figure 5-1**, to be detailed in the methods).

To test the viability and developmental sensitivity of the metaphor reference task, Experiment 1 tests two relatively distant age groups of 7- and 10-year-olds. We expect to find the older participants’ performance with the task to reflect greater competence than the younger ones’. Likewise, we expect that the picture choice to be slower for correct metaphoric choices when compared to correct synonymic choices.

The other goal of the current study concerns deepening our knowledge about metaphor comprehension during the school-age years. This leads to Experiment 2, which investigates a larger sample of children (between the ages of 6 and 10) as they carry out the metaphor-reference task. Besides aiming to confirm the developmental findings from Experiment 1 in finer detail, it will also assess children’s formal language skills and TOM, in order to gauge the contributions that these two factors make on metaphor comprehension. Previous studies focused on the importance



**Figure 5-1:** Illustration of a vignette pair structure (translated from French to English) depicting what a participant sees (top) and hears (bottom)

of one or the other of those potential predictors. Consider Lecce et al., (2019) or Tonini et al. (2023), who use language measures as a control variable, but focus their investigation on the role of TOM. The roles of these two predictors have yet to be studied jointly in a single study.

To underline what else is novel in this work, we point out two other features of the task's implementation. One is that the task was prepared on a tablet which allows young participants to work at their own pace in the comfort of their classroom. The other is that, while the linguistic abilities of children in Experiment 2 will be assessed through a standard measure, the *BILO-3C Oral comprehension* task (Khomsi et al., 2007), the TOM measure will be determined by another tablet-based adaptation, the Langdon & Coltheart [1999]'s Picture Sequencing Task (PST), which was validated in Petit et al. (under review, see *Annexe B*). As the title of the tablet-PST suggests, it requires participants to put pictures in a thematic order. One class of items in the tablet-PST requires participants to represent a false belief (e.g., a character who mistakenly assumes that their dog ate a pie that was cooling when in fact it was

stolen by someone through the kitchen's window). Prior studies show that such false belief representations are harder to sequence than controls. Importantly, this task relies on minimally verbal material, which rules out the possibility of a language confound. Below, we describe the Experiments, as well as its adjacent tasks, in greater detail.

## Experiment 1

The main goal of this Experiment is to test whether a novel variation of the metaphoric reference task provides developmental effects that are consistent with prior studies, in which children's metaphor comprehension increases with age. In the current Experiment, participants are aurally presented short vignettes that include a character and a set of events. At the end of the vignette, an earlier-mentioned element is referred to in a metaphoric or synonymic way. To test participants' comprehension, they are then presented with a picture-choice task that, for the metaphor condition, juxtaposes a picture showing the intended figurative referent with one showing a

literal representation of the reference. Importantly, this task allows children to work independently via a tablet that collects accuracy data as well as latencies.

## Method

### Participants

Eighty-nine children took part in this study. All were native French speakers recruited from a local private school. None had neurodevelopmental disorders (learning, autism spectrum, attention-deficit, or language-related disorder) nor any sensory (visual or auditory) or motor disorder that would prevent them from using a tablet, as declared by parents. The study included two age groups: young children from 1<sup>st</sup> and 2<sup>nd</sup> grade (N = 48, 48% girls, mean age = 7;0, SD = 0;7) and older children from 4<sup>th</sup> and 5<sup>th</sup> grade (N = 41, 32% girls, mean age = 10;0, SD = 0;7). Ethnicity is not reported since the collection of such information is forbidden by French laws. All children were free to participate, and their parents provided informed non-opposition, in accordance with local research ethics rules.

### Materials

We developed a referential task inspired by Noveck et al.'s (2001) and Van Herwegen et al.'s (2013) paradigms. Each item consists of an orally-presented vignette in French, uttered by the first author and split into three parts (see **Figure 5-1**). In the first part of the vignette, two to three sentences provide background including mention of a noun phrase that will be referred back to in the last part of the vignette in the target sentence. It is this later reference that will be either metaphoric or synonymic.

Each vignette is introduced with two slightly different contexts. To put it another way, the target reference is designed to be either a metaphoric or synonymic reference

with respect to an earlier mentioned element that depends on the initial context (e.g., the reference *fountain* in Figure 1 can be used to refer back to a *crying baby* in the metaphoric condition or to *water jets* in the synonymic condition). The metaphoric target expressions were designed to be novel. Note that, unlike in Noveck et al. (2001), which used different target words to refer to a singular previously-mentioned element, in the current task the *initial referents* vary across two slightly different contexts before setting up a singular target reference.

Another general feature of the paradigm includes the display of a picture that is thematically related to the vignette. This is done in order to engage the child's attention. Technically speaking, the task is self-paced; to advance the story, participants are required to press a speaker button in order to hear the next part of the story. Participants can listen to a segment as often as they pleased through a 'repeat' button.

After the vignette's presentation is completed, a question about the target reference is uttered by a female voice (for example, in **Figure 5-1**, she asks "*What will Mary take care of?*"). Immediately after the end of the question, four pictorial options are displayed on the screen. From the point of view of the metaphoric reference condition in **Figure 5-1**, the picture options reflect a) a representation of the metaphoric reference that harks back to the earlier referent (the crying baby), b) a representation of the literal meaning of the target reference (a picture of a fountain), c) a distractor that is distantly related to the story content (a sandbox which could be found in a park), and; d) an "I don't know" symbol. Both the participants' picture selection and their selection responses times are automatically recorded by the tablet.

For expository purposes, from here on when we refer to the two versions of each

vignette, we will call them *metaphoric* and *synonymic* because the reference in the target sentence refers back to an earlier mentioned element in one of those two ways. When we consider the picture options that are available, we will refer to the intended meaning of the target reference. Thus, after a vignette makes a *metaphoric* reference, an accurate response ultimately involves choosing a picture that depicts the original referent (the *crying baby*). On the other hand, after a vignette makes a *synonymic* reference, an accurate response involves choosing a picture that corresponds to the literal meaning of the target reference. The upshot of this approach is that in the wake of a metaphoric reference, we expect the literal representation of the target reference to have some appeal for participants but at decreasing rates with age. In contrast, note that the correct response in the synonymic condition is expressed through a representation of the literal target reference only (the fountain). For this condition, both of the two remaining pictorial options are not related to the story.

Twelve vignettes were designed so that each could produce two versions, one that provides for a metaphoric reference and one a synonymic one. Efforts were made to limit the linguistic demands of the task, with simple sentences and high frequency words. Stories are matched (see **Table S1.1** in *supplementary materials*) for (1) length, (2) readability, (3) mean frequency of content words (nouns, verbs, adjectives, adverbs) in an oral French corpus, and (4) frequency of referents in school textbooks.

As far as what an individual participant sees, each vignette was presented only once (in one of its two versions). The experiment was designed to distribute equal amounts of metaphoric and synonymic versions; so, four lists of items were prepared. In each, six vignettes were ultimately presented in their metaphoric version and six in their

synonymic one. All told, each vignette theme appeared in their metaphoric version in two lists and in their synonymic version in the two other lists, so that each version of each vignette appeared together with two different sets of stories. The presentation order of the 12 vignettes across the four lists (once they were assigned their metaphoric or synonymic status) was based on a randomized procedure, but to control for potential *trial-order* effects, each list order was also reversed, so that ultimately 8 different lists were distributed.

To familiarize participants with the task and response format, an additional training vignette, involving a synonymic reference, was added prior to the experimental session. This had to be successfully completed in order for the proper task to begin. Instructions were automatized and included in the app. The task lasted 8-10 minutes.

## **Procedure**

Children were tested in their usual chairs in their customary classroom, with 15 to 25 children completing the tasks simultaneously. Each child was provided with a tablet and a headset, and each was systematically checked, along with the accompanying app, before testing began. Each tablet had one of the eight lists randomly assigned to it. Each child was instructed that they would have to listen to each of the stories carefully, so as to be able to later answer questions. They were then instructed to launch the app, which began with a demo of how the response system worked along with the practice item. During the practice phase, the experimenter remained available to answer any questions or to help if necessary. Feedback on the practice item was automatically provided by the app, after which children were prompted to carry out their task at their own pace. It was made clear that no feedback or help could be provided by the experimenter during the testing phase. The metaphor task

was proposed in tandem with another 5-minute-long experimental task (that is not reported here) and children were allowed to begin with the task of their choice.

### Analysis

Accuracy and responses times to the metaphor task were analyzed with (generalized) mixed effect models, fitted in R (R Core Team, 2022) with the *lme4* package (Bates et al., 2015). Likelihood ratio tests were used to assess fixed effects and post-hoc contrasts were computed with the *emmeans* package (Lenth, 2022). Response time analyses were run on log-transformed data of correct answers only, after removing outliers presenting a deviation of more than 2 *SD* from each age group's mean, within each condition. Sum contrasts were used for all independent variables.

### Transparency and openness

All data and analysis code are available at [https://osf.io/myr2c/?view\\_only=7c8aeebe9837458ea5eb1c365bf20ccd](https://osf.io/myr2c/?view_only=7c8aeebe9837458ea5eb1c365bf20ccd). This experiment design and its analysis were not pre-registered.

## Results

Before beginning our developmental analyses, we first verified that the lists, which were randomly attributed to children, were evenly distributed across age; this was statistically confirmed ( $\chi^2(3) = 0.5, p = .66$ ).

### Accuracy

Here, we turn to rates of accuracy on participants' picture selection. We constructed a generalized model using age group and reference condition as fixed effects, as well as their interaction, with random intercepts for items and

participants. This revealed a main effect of condition ( $\beta = 3.4, SE = 0.25, p < .001$ ) and an interaction of condition by group ( $\beta = 3.2, SE = 0.45, p < .001$ ). As far as the metaphoric reference task is concerned, we observed that participants provided post-reference picture-choice accuracy rates of 43% overall along with a clear developmental pattern. Twenty-five percent of the 7-year-olds' choices in the metaphoric condition were accurate, while 63% of the older children's were ( $\beta = -2.0, SE = 0.28, p < .001$ ). Most errors among the metaphoric items were due to children's attraction to the pictures depicting a literal representation of the target reference (these accounted for 97% of errors, see **Figure S1.1** in *supplementary materials*). The two other response propositions were very rarely chosen. Contexts designed to prompt a synonymic reference mostly induced correct picture choices (at a rate of 93%). Interestingly, the contrast between age-groups was significant for this condition too but in a direction opposite of what one might expect. While practically at ceiling, the 7-year-olds committed slightly fewer errors (4%) than the 10-year-olds (10%;  $\beta = 1.2, SE = 0.4, t = 2.7, p < .01$ ). It will be seen that this result – where the 10-year-olds become wary of alternative readings in the synonymic condition – re-emerges in other findings.

To confirm that results from the task were not adversely affected by specific instantiations of the vignettes, we added *list* to the model as a fixed effect, as well as its interaction with the two other factors. This did not significantly improve the model ( $\chi^2(12) = 19.6, p = .08$ ) and, importantly, it did not alter the pattern described above. Likewise, we examined results by vignette (see **Table S1.2** in *supplementary materials*). Item analyses confirmed that for each vignette, picture choice was determined by the preceding context. That is, accurate picture choices reflect post-

metaphoric as opposed to post-synonymic reference making. More importantly, after a metaphoric reference, 10-year-olds were more likely than 7-year-olds to choose pictures accurately (as intended by the vignette's context).

For exploratory purposes, we then added *item position* as a fixed effect to the initial model, as well as its interaction with the two other fixed effects. This significantly improved the model's fit ( $\chi^2(4) = 46, p < .001$ , see **Figure S1.4** in *supplementary materials*), revealing no presentation order effect in the younger age group (all  $p > .05$ ), but clear effects in the older group. That is, ten-year-olds' performance improved over the course of the task with respect to the metaphoric reference condition ( $\beta = 0.22, SE = 0.05, z.ratio = 4.7, p < .001$ ) but they became slightly but progressively weaker in terms of accuracy in the synonymic condition ( $\beta = -0.31, SE = 0.08, z.ratio = -3.8, p < .001$ ). This is in line with the finding concerning 10-year-olds performance in the synonymic condition described above, where they appear to make more errors than the 7-year-olds. We will provide an account for this finding in the Discussion section.

### Response times

Children's latencies while making their picture choices were analyzed with a linear mixed effect model on the correct responses. Per our pre-analytic procedure, we filtered out data that were excluded as outliers (representing 4% of the data), leaving a total of 690 data points. We constructed a linear model using the same structure as for accuracy: age group and reference condition as fixed effects, as well as their interaction, with random intercepts for items and participants.

This model revealed a main effect of age ( $\beta = 0.19, SE = 0.06, t = 3.3, p < .01$ ), as older children (mean  $RT = 2966$  ms,  $SD = 1534$ )

were faster than younger children (mean  $RT = 3172$  ms,  $SD = 1263$ ). There was also a main effect of reference condition with a slowdown induced by metaphoric contexts ( $\beta = -0.22, SE = 0.03, t = -7.0, p < .001$ ), with no interaction ( $\beta = -0.07, SE = 0.06, t = -1.1, p = .26$ ). Adding *list* as a fixed factor improved the model's fit ( $\chi^2(8) = 21.1, p < .01$ ). While certain lists appeared more sensitive to effects of reference than others and while certain other lists were more sensitive to age effects, the global pattern concerning age effects and reference-condition were not unduly affected (see **Figure S1.3** in *supplementary materials*).

Finally, we added *item position* in the task as a supplementary fixed effect to the initial model (**Figure S1.4** in *supplementary materials*), which also significantly improved its fit ( $\chi^2(4) = 22.7, p < .001$ ). Similar to our findings on accuracy, no pattern was evident among the 7-year-olds (all  $p > .05$ ). In contrast, the older children tended to respond faster in correctly choosing their picture as the task progressed after making metaphoric references ( $\beta = -0.02, SE = 0.008, t = -2.8, p < .01$ ) and to respond more *slowly* in correctly choosing their picture as the task progressed after making synonymic references ( $\beta = 0.02, SE = 0.007, t = 3.5, p < .001$ ).

## Discussion

In Experiment 1, we tested a new tablet-based metaphor reference task on 89 typically developing children. Participants were presented with vignettes that ultimately led to the presentation of a target reference that could, depending on the prior context, be metaphoric or synonymic. Our main goal was to determine that this new task was accessible to children while producing results that resonate with prior findings.

Based on the results of the current Experiment, we can draw the following five conclusions. First, given children's performance with the control items (the synonymic-reference condition), one can be confident that young participants are carrying out the task competently. Error rates were low overall for both age groups. Second, the data show that children were sensitive to preceding context, providing evidence for the task's face validity. Vignettes that prompted the listener to make a metaphoric reference led to more equivocality among all the children. That is, the vignettes that were designed to prompt metaphoric references were also prone to errors (overall, depictions of a literal representation of a metaphor target reference were chosen roughly half the time). Third, and most importantly to our study, rates of accurate (intended) metaphoric choices increase with age: the 10-year-olds were more likely than the 7-year-olds to choose pictures that show that they made metaphoric references. These observations also applied to vignettes taken individually. Fourth, the post-metaphoric-reference picture choices come with a cost in terms of response times when compared to post-synonymic-reference picture choices that follow from the same target expressions. This makes sense for the simple reason that, generally speaking, metaphoric picture-choices were juxtaposed with tempting literal depictions of the target reference. Fifth, the addition of finer exploratory analyses – such as those that add list effects or trial effects – only enhance the main results. Overall, these results confirm the task's viability as well as its ability to lay bare a developmental sensitivity of metaphor comprehension over the school age years.

The 10-year-olds in particular produced interesting outcomes. Among other things, we observed that performance changed over the course of the task for this age group.

Notably, the 10-year-olds' picture choices made in the wake of the metaphoric-reference target sentence were carried out with both greater accuracy and decreased latency over the course of the experimental session. Interestingly, we also observed that there was a slight negative effect with respect to their picture choices in the wake of the synonymic target-reference. The 10-year-olds were slightly, but significantly, more likely to consider alternatives even to the synonymic reference; likewise, the 10-year-olds took longer to correctly choose the picture that corresponds to the literal representation of the target reference, as the task wore on. We surmise that, based on their experience with the task and its multiple metaphoric items, children progressively became aware of dual meanings in the task and that this prompted them to be more discerning, even among the vignettes that were designed to have a univocal interpretation. Indeed, errors in the synonymic condition that augmented in the course of the task reflected metaphorical interpretations of the reference. Generally speaking, this ability to appreciate an alternative meaning captures well how 10-year-olds differ from their younger cohorts.

## Experiment 2

Experiment 1's results allow us to conclude that the metaphor referential task here provides the kind of outcomes one would expect based on prior studies. That is, metaphor-reference abilities appear to become more adult-like with age. Likewise, one generally finds metaphor-related slowdowns, which also resonate with the previous literature. In the present case, the results are based on the way participants make their choice of pictures (and not, for example, on reading times).

In Experiment 2, the same task is used to further analyze the developmental effect

with the additional aim of determining the extent to which two different growing abilities – linguistic competence and TOM – contribute to children’s metaphoric performance. That is, we investigate interindividual differences in children’s metaphor comprehension in terms of age, like in Experiment 1, and this time while also considering other cognitive abilities, which can be used as predictors in analyses. In order to provide this sort of assessment, the task is presented in the context of a standardized linguistic-pragmatic assessment battery. This is why it is presented in a single order and while interdigitated with other pragmatic tasks. In other words, the metaphor reference task here is similar to the one in the prior experiment, except that it is embedded in a larger and longer exercise, which includes in effect filler items. Assuming that we replicate the findings from Experiment 1, it would only increase the metaphor reference task’s validity.

## Method

### Participants

Two hundred and forty-eight children took part in this study. Children were between 6;0 to 10;11 years old (mean = 8;4 years old,  $SD = 1;5$ ) and were homogeneously distributed across age groups (see **Table 5-1**). As in Experiment 1, all participants

regularly attended an elementary school (*école primaire*) and were native speakers of French. Both parents and children provided informed consent. As declared by parents, children had no neurodevelopmental disorder (learning, autism spectrum, attention-deficit, or language-related disorder), nor any sensory (visual or auditory) or motor disorders that would prevent them from using a tablet.

Language in general is much associated with children’s socio-economic environment (see, e.g., Di Sante & Potvin, 2022). To provide social balance, children were thus recruited from two different French schools, one private school in an urban economically privileged area (School A) and one public school situated in a rural and less economically privileged region (School B). Moreover, we collected, via parental questionnaires, an estimation of family income through the Family Affluence Scale (FAS, Currie et al., 2008), which is based on relatively non-intrusive questions and has proven to be a good indicator of family wealth across countries (Boyce et al., 2006; Currie et al., 2008). We also asked parents to provide their education levels, which were coded on a scale ranging from 0 (no diploma) to 7 (PhD). For each child, parents’ mean education level is used when information about both parents is available; when data from only one parent is available (6 % of children), the score of that one parent is reported. The FAS and the parental

**Table 5-1:** Sample characteristics by age groups

	6 YO	7 YO	8 YO	9 YO	10 YO	Whole sample	Age group difference
N	53	59	47	41	48	248	/
Proportion female	43%	54%	45%	56%	58%	51 %	$\chi^2(4) = 3.7,$ $p = .45$
Proportion from school A	54%	54%	36%	58%	50%	51%	$\chi^2(4) = 4.6,$ $p = .33$
Mean SES Index (SD) <sup>1</sup>	59 (15)	59 (18)	57 (16)	62 (17)	56 (16)	58 (16)	$F(4,244) = .85,$ $p = .50$

<sup>1</sup> Data were missing for 3 participants who were removed from these analyses.

education level are standardized then averaged to provide a socio-economic status composite (SES) varying from 0 to 100.

This study was authorized by the two schools where children were included and tested. It was part of a project which also required the inclusion of participants with neurodevelopmental disorders (not reported here), and as such received ethics approval from the [name of the board and ID RCB, *masked for review*].

## Materials

### Metaphor task

This Experiment employed a unique version of the task by selecting one of the lists from Experiment 1. The advantage of preparing the task in this way is that potential differences between two children cannot be attributed to list or presentation order effects. Moreover, this approach will allow us to consider the data collected as normative for individual subjects, which will subsequently permit us to make clinical applications while also extending our research on interindividual differences to atypical development.

As part of the unique version employed here, we mixed the 12 items from the first list of Experiment 1 with 22 filler items of similar structure and response format which involved other language and cognitive abilities such as memory, irony or conditional reasoning. These are not of interest to the present paper and will not be discussed further. For this version, the task is preceded by 3 training items to familiarize children with the response format, which relied on literal material only and provided the participant with automatic feedback. To sustain engagement and attention for this longer version of the task, two breaks were included at approximately a third of the way through and then at three-quarters of the way through, at which points

participants viewed 30 seconds of non-verbal video-clips of funny animals. The task lasted approximately 20 minutes.

As was the case for Experiment 1, we verified that in the selected list, the vignettes that set up the metaphoric and synonymic references remained matched for (1) length, (2) readability, (3) mean frequency of stories content words (nouns, verbs, adjectives, adverbs) in an oral French corpus, and frequency of both (4) referent and (5) target nouns in school textbooks (see **Table S2.1** in *supplementary materials*).

### Parental appreciation of pragmatic abilities

To provide external validation for the metaphor task, we used a subset of the validated parental questionnaire of children's communicative abilities developed by Bishop (2003), the French adaptation of the Children's Communication Checklist-2 (Vézina et al., 2013). In this questionnaire, parents are asked to judge the frequency at which they observe different communication-related behaviors in their child. We extracted the items from the "Use of context" subscale which targets context-dependent uses of language such as figurative language, including metaphors (e.g., "How often is your child over-literal, sometimes with (unintentionally) humorous results?"). As a control, we used the items from the "Semantics" subscale, which targets non-pragmatic linguistic behaviors, such as word finding (e.g., "How often does your child make false starts, and appears to grope for the right words?"). Each subscale is composed of 7 items and provides a total score ranging from 0 to 21, with higher values being associated with higher communication difficulties.

### Grammar task

Grammar reception was assessed with the *BILO-3C Oral comprehension* task

(Khomsî et al., 2007). This is a computerized French-validated picture-matching task targeting sentences with different morpho-syntactic phenomena such as verbal inflexions of number and tense, passive structures, object pronouns, direct and indirect relative clauses. For each item, four pictures are displayed on a screen, and children are asked to select the picture that best-matches an aurally presented sentence. For example, to assess passive structures, children are aurally presented a sentence, such as *Marie est poussée par Pierre* (*Marie is pushed by Pierre*), while being shown four pictures: i) a boy pushing a girl on a bicycle (expected answer), ii) a girl pushing a boy on a bicycle, iii) a boy watching a girl on a bicycle, iv) a boy and a girl pushing a stroller). It includes 2 training items and 27 test items, and provides a total score ranging from 0 to 27.

#### Theory of mind task

TOM was assessed with a tablet-based version of the Picture Sequencing Task (PST), an experimental task designed by Langdon & Coltheart (1999). In this minimally-verbal task, participants are asked to sequence four pictures. Following guidelines developed by Langdon & Coltheart (1999), participants receive 6 points for properly positioning all of the pictures of each sequence (2 points for each of the first and last picture, and 1 point for each of the intermediate pictures). Following Rajkumar et al. (2008), we used a subset of 12 items from Langdon & Coltheart's (1999) original material, broken down evenly across three conditions. That is, four sequences involve mechanical causalities (e.g., consider pictures showing how a speeding truck's vibrations prompt a boulder to roll down a hill), four were based on social scripts (e.g., consider a sequence that represents meeting a friend for a coffee) and four critical sequences involve false beliefs (as described above, these involve

detecting a false attribution). There were also two training items. Scores for the first two control categories are averaged to provide a General Sequencing Abilities (GSA) index, which has proven to be strongly correlated with IQ among typically developing school-aged children (Rajkumar et al., 2008). This measure is used as a control measure while the average of the false-belief scores constitutes the TOM index. Both indexes vary from 0 to 6. The PST has been used to study TOM skills in various populations, including high vs. low schizotypal healthy participants in the original study (Langdon & Coltheart, 1999), children and adults with William's syndrome (Porter et al., 2008) as well as adults with bipolar disorder (Van Rheeën & Rossell, 2013). The tablet-PST's overall Cronbach's alpha was .76 in our sample (for more information, see Petit et al. [under review, see [Annexe B](#)]).

#### Procedure

Parental consent forms and questionnaires (inclusion criteria check, SES, CCC-2) were collected before testing. As in Experiment 1, assessments took place in the children's usual classrooms. Children were first tested with a portion containing the metaphor task; this took roughly 20 to 30 minutes. TOM and language assessments took place in a second session on a different day. This second day of testing lasted roughly 20 to 30 minutes as well.

#### Analysis

Before analyzing the data, we took two preliminary steps. First, as this study is targeting typical development only, we removed those participants whose rates of correct responding were 2.5 SD below their age group's mean in the grammar or TOM task. Second, for response time analyses, a) the data were log-transformed and b) incorrect responses and outliers were removed based on the procedure used in

Experiment 1. Participants with missing data for SES were also excluded from the analysis using this variable. All analyses were performed in *R* (R Core Team, 2022).

We begin by testing the structural and external validity of the metaphor-reference task. This is first done by assessing its 2-factor structure (metaphoric vs synonymic reference) with a confirmatory factor analysis (CFA), using the *lavaan* package (Rosseel, 2012). Then, we examined Cronbach's alpha for each of these 2 conditions. To determine external validity, we computed correlations between accuracy in the critical condition of the metaphor reference task with i) the CCC-2 "Use of context" subscale, to provide convergent validity (where a correlation is expected), and ii) the CCC-2 "Semantics" subscale to provide divergent validity (where no correlation is expected).

Then, we fitted mixed effects models, i.e., generalized mixed effect models (for accuracy) and linear mixed effects models (for response times). These were fitted with the *lme4* package (Bates et al., 2015). In order to address our research concerns, we followed the following two steps. First, we fitted hierarchical models that include reference condition as well as single, quadratic and then cubic terms for age as fixed effects, as well as their interactions with condition. Random structure included random intercepts for participants and items. This should allow us to describe the precise evolution of performances with age. Second, with the aim of assessing the role of the different cognitive predictors on the DV, we dropped the higher order terms for age, to prevent computation as well as interpretation issues. Then we added the predictors of interest as fixed effects (grammar and TOM, as well as the GSA, to control for sequencing abilities) and allowed for 3-way interactions of these predictors with age and condition. In that way we could describe each predictor's

effect across reference conditions and at different ages. SES was also added as a control covariate. The final model formula was thus  $DV \sim SES + age * condition * (Grammar + GSA + TOM) + (1|participant) + (1|item)$ .

In each model, age was entered in years as a continuous predictor (this follows Royston et al. [2006] who argue that continuous predictors should not be dichotomized). For descriptive purposes, and because interactions involving multiple continuous predictors are complicated to examine, effects at different ages were described via post-hoc contrasts run from the fitted models at the values 6, 7, 8, 9 and 10 years of age (as could have been done for any particular value) and through visualizations. This should not be taken to imply that we used age-groups as a categorical variable in the analysis. Sequential difference coding was used as a contrast for age and sum contrast for reference condition. All continuous predictors were centered on their mean with the *scale()* function. Models' assumptions were checked before reporting their estimates. Single effects were assessed with likelihood ratio tests. All post-hoc contrasts were performed with the *emmeans* package (Lenth, 2022) while adjusting for multiple comparisons.

### Transparency and openness

All data and analysis code are available at [https://osf.io/myr2c/?view\\_only=7c8aebe9837458ea5eb1c365bf20ccd](https://osf.io/myr2c/?view_only=7c8aebe9837458ea5eb1c365bf20ccd). This experiment design and its analysis were not pre-registered.

## Results

In the sample, 10 participants out of 248 (4%) scored below 2.5 *SD* from their age-group mean on the control tasks and were thus excluded. One participant did not

**Table 5-2:** Final sample (N=236) results on the control tasks by age groups, and correlations with age

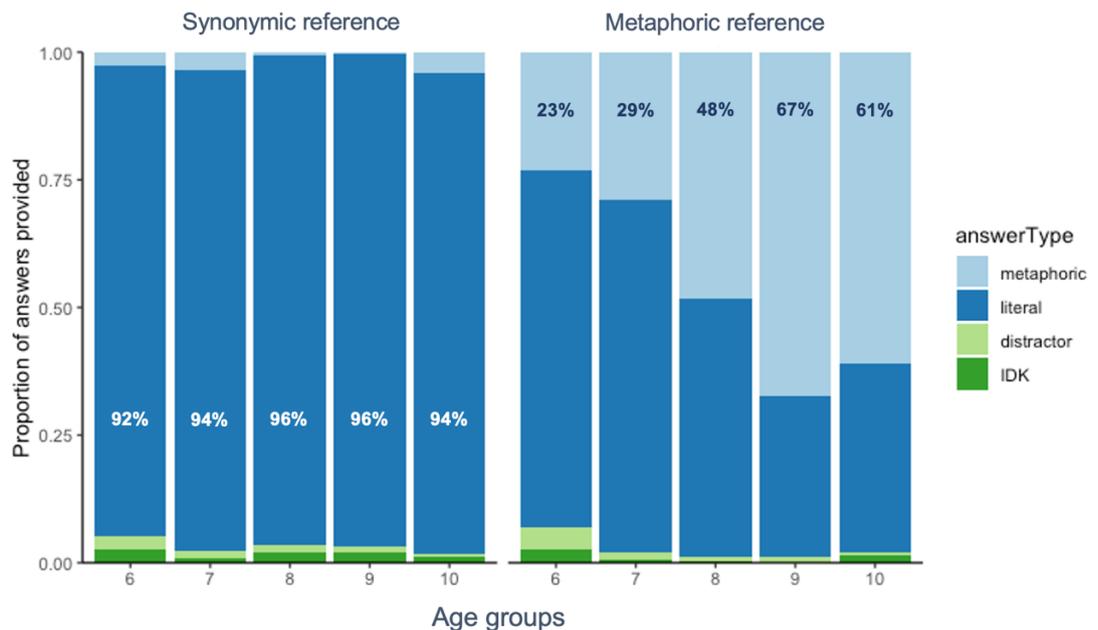
	Groups	6 YO N = 49	7 YO N = 55	8 YO N = 45	9 YO N = 41	10 YO N = 46	Correlation with age
Grammar	Mean	13.2	14.5	16.4	17.9	18.3	$r = .51$
	(SD)	(3.53)	(2.98)	(3.71)	(2.92)	(3.50)	$p < .001$
TOM	Mean	2.90	3.43	4.13	4.54	4.02	$r = .37$
	(SD)	(1.01)	(1.17)	(1.24)	(1.10)	(1.26)	$p < .001$
GSA	Mean	4.44	4.77	4.94	5.47	5.27	$r = .36$
	(SD)	(1.19)	(0.82)	(0.74)	(0.56)	(0.80)	$p < .001$
CCC-2 Semantics	Mean	6.6	6.0	6.3	6.2	6.1	$r = -.0$
	(SD)	1.7	1.6	1.7	1.7	1.4	$p = .23$
CCC-2 Context	Mean	5.3	4.9	5.1	3.5	3.9	$r = -.21$
	(SD)	2.7	2.6	3.4	2.5	2.5	$p < .01$

complete the control task and was also excluded, leaving a remaining sample of 237 children for the analysis. For descriptive purposes, **Table 5-2** presents the distribution of children across age-groups and their results on control tasks.

### Task validation

The confirmatory factor analysis clearly confirmed the bifactorial structure of the metaphor task, one that contrasts synonymic and metaphoric items ( $CFI = .98$ ,  $TLI = .98$ ,  $RMSEA = .03$ ), with latent factors sharing no variance ( $\beta = 0.001$ ,  $SE = 0.001$ ,  $p = .49$ ). The internal consistency of

the metaphorical condition proved to be excellent without revealing redundancy in the items ( $\alpha = .81$ , 95%  $CI = [.77-.84]$ ); meanwhile, the synonymic condition appeared to provide less consistency ( $\alpha = .50$ , 95%  $CI = [.41-.60]$ ), which could be accounted for by the ceiling effect observed in this condition (see below). Due to this ceiling effect, convergent and divergent validity was assessed only for the metaphor condition. For the 225 children whose parents filled the entire questionnaire, the individual mean accuracy rate for metaphoric reference was significantly correlated with the parental report of



**Figure 5-2:** Proportion rates of each answer for the different conditions (based on raw data), across the different age groups, with accuracy for each age group in each condition

PREDICTORS	ACCURACY			LOG(RT)		
	Odds Ratios	95% CI	p	Estimates	95% CI	P
(INTERCEPT)	5.10	3.49 - 7.47	<0.001	8.25	8.03 - 8.47	<0.001
SES	1.20	0.98 - 1.47	0.072	<b>-0.04</b>	<b>-0.07 - -0.00</b>	<b>0.023</b>
AGE	<b>1.50</b>	<b>1.16 - 1.94</b>	<b>0.002</b>	<b>-0.04</b>	<b>-0.06 - -0.01</b>	<b>0.008</b>
CONDITION	<b>0.01</b>	<b>0.01 - 0.03</b>	<b>&lt;0.001</b>	<b>0.34</b>	<b>0.05 - 0.62</b>	<b>0.022</b>
GRAM	1.16	0.88 - 1.52	0.304	-0.06	-0.27 - 0.14	0.542
GSA	1.12	0.86 - 1.46	0.409	-0.08	-0.26 - 0.10	0.387
TOM	1.00	0.78 - 1.28	0.989	<b>-0.22</b>	<b>-0.44 - -0.01</b>	<b>0.043</b>
AGE * CONDITION	<b>1.86</b>	<b>1.28 - 2.71</b>	<b>0.001</b>	-0.02	-0.05 - 0.01	0.231
AGE * GRAM	1.11	0.87 - 1.42	0.380	0.01	-0.02 - 0.03	0.472
AGE * GSA	0.91	0.72 - 1.15	0.441	0.01	-0.01 - 0.03	0.428
AGE * TOM	0.99	0.78 - 1.27	0.966	0.02	-0.00 - 0.05	0.067
CONDITION * GRAM	1.69	1.12 - 2.54	0.012	0.24	-0.02 - 0.50	0.074
CONDITION * GSA	<b>0.66</b>	<b>0.45 - 0.97</b>	<b>0.033</b>	-0.20	-0.44 - 0.04	0.103
CONDITION * TOM	<b>1.51</b>	<b>1.04 - 2.20</b>	<b>0.029</b>	<b>-0.36</b>	<b>-0.64 - -0.07</b>	<b>0.013</b>
(AGE * CONDITION) * GRAM	<b>2.31</b>	<b>1.62 - 3.28</b>	<b>&lt;0.001</b>	<b>-0.03</b>	<b>-0.06 - -0.00</b>	<b>0.034</b>
(AGE * CONDITION) * GSA	1.14	0.81 - 1.60	0.453	0.02	-0.01 - 0.05	0.168
(AGE * CONDITION) * TOM	<b>0.71</b>	<b>0.50 - 1.00</b>	<b>0.05</b>	<b>0.04</b>	<b>0.01 - 0.07</b>	<b>0.018</b>
<b>RANDOM EFFECTS</b>						
$\Sigma^2$	3.29			0.12		
T <sub>00</sub> PARTICIPANT	1.19			0.03		
T <sub>00</sub> ITEM	0.26			0.01		
ICC	0.31			0.25		
OBSERVATIONS	2808			1861		
MARGINAL R <sup>2</sup> / CONDITIONAL R <sup>2</sup>	0.493 / 0.648			0.071 / 0.301		

**Table 5-3:** Outcomes of the generalized mixed effects model explaining accuracy in the task (left part, orange) and of the linear mixed effects model explaining log-transformed response times to the picture selection portion of the task (right part, green)

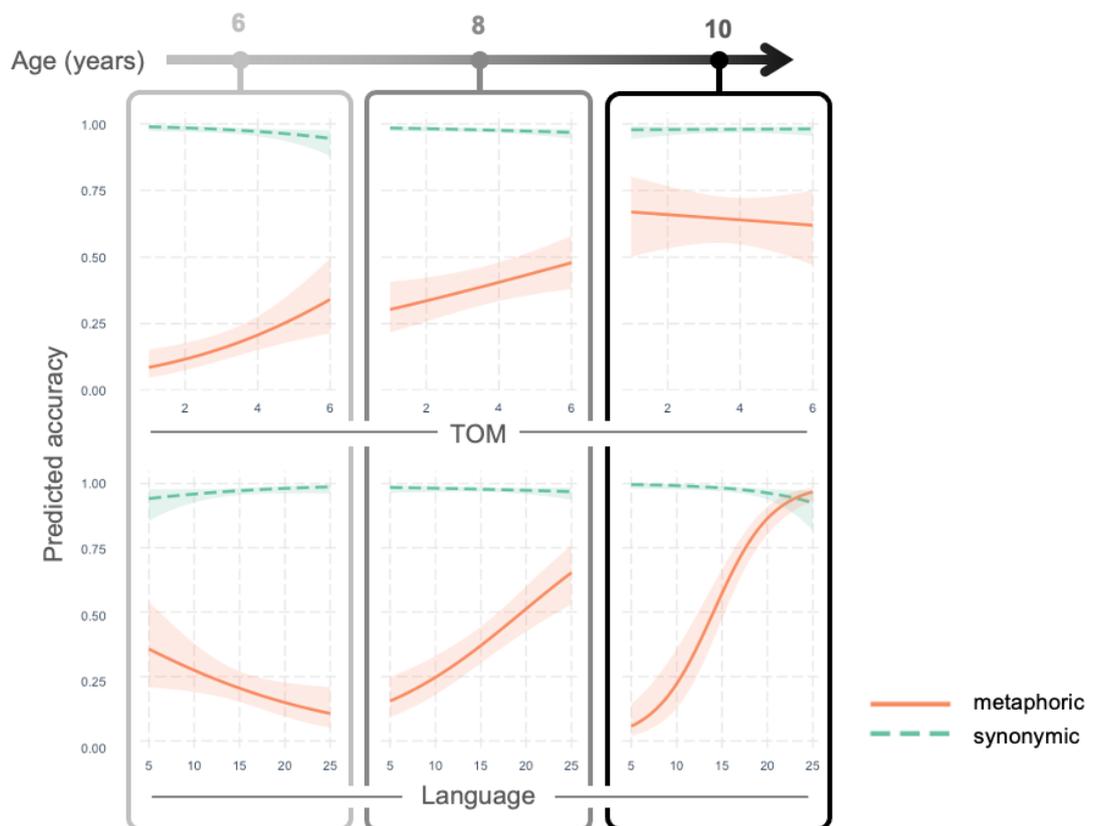
pragmatic abilities (CCC-2 “Use of context”  $r = -.21$ ,  $p < .01$ ) but not with semantic abilities (CCC-2 “Semantics”,  $r = .02$ ,  $p = .81$ ). Finally, the results of experiment 2 were very consistent to those of experiment 1: the seven- and 10-year-olds accurately choose the metaphoric representations at rates of 29% and 61%, respectively in this Experiment, contra 25% and 63% in Experiment 1.

### Accuracy

Accuracy in the synonymic condition was very high overall (95%), indicating that participants correctly chose the intended target (a picture reflecting a literal representation of the reference) here. As expected, and as shown in **Figure 5-2**, rates of correct picture choices in the metaphoric-reference condition were lower (44%). Also, as expected, most errors resulted from the selection of the pictorial representation

of the reference taken literally (this accounted for 95% of errors).

The first-step analysis revealed that the cubic model outperformed the linear model ( $\chi^2(4) = 16.9$ ,  $p < .01$ ). This model confirmed that participants were much more likely to succeed the synonymic-reference items than they were in the metaphoric-reference items ( $OR = 4.1$ ,  $SE = 0.38$ ,  $p < .0001$ ). Post-hoc contrasts revealed no age effect in the synonymic condition. As for the metaphoric-reference condition, the 6- and 7-years-olds prompted the lowest accuracy rates, the 8-year-olds scored higher than the youngest participants ( $\beta = 1.2$ ,  $SE = 0.2$ ,  $p < .0001$ ), and the 9-year-olds outperformed the 8-year-olds ( $\beta = 1.2$ ,  $SE = 0.2$ ,  $p < .01$ ). The 10-year-olds’ rates of accuracy were ultimately comparable to those of the 9-year-olds ( $p > .05$ ).



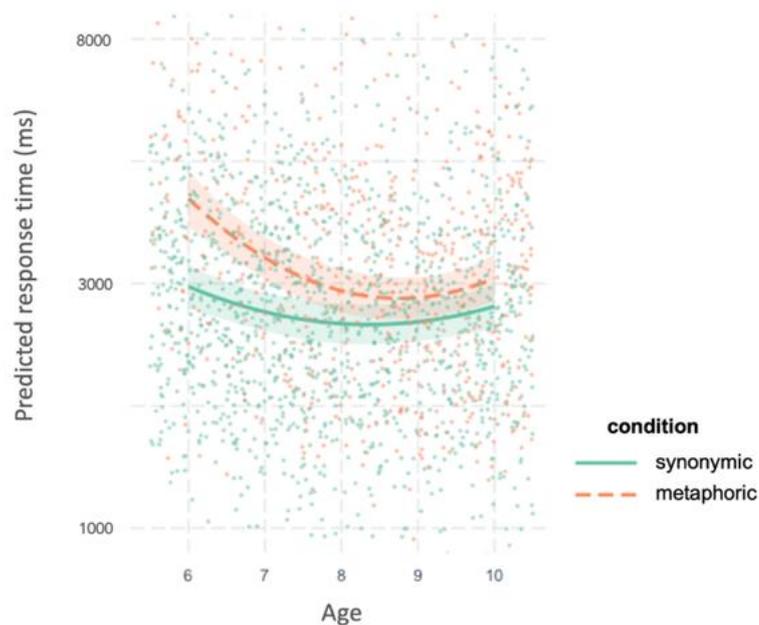
**Figure 5-3:** Predicted accuracy in the different conditions, as a function of TOM (*x*-axis, above) and grammar (*x*-axis, below), at 3 different ages (6-, 8- and 10-year-olds) (predictions were backtransformed from scaled variables)

In a second step, we dropped the cubic and quadratic terms for age and added the additional predictors (see model's output in **Table 5-3**). This model revealed that accuracy was overall more predicted by both TOM and grammar in the metaphoric than in the synonymic, reference condition. These variables and the reference condition were involved in 3-way interactions with age: TOM appears to have a positive effect on picture choice accuracy in the metaphoric reference, as opposed to the synonymic reference condition, but the influence of TOM decreased with age. As for the role of grammar, one sees the reverse pattern, i.e., it reveals having an *increasing* influence on metaphor accuracy with age, compared to the synonymic condition. Of the two, grammar appears to be more impactful on the three-way interaction than TOM is.

The predicted marginal effects of TOM and language on accuracy in the reference task,

as a function of age, is plotted in **Figure 5-3**. Post-hoc contrasts (see **Tables S2.2** and **S2.3** in *supplementary materials*) revealed that higher scores on TOM predicted a higher accuracy on the metaphoric reference task compared to the synonymic controls at 6, 7 and 8 years of age, but not at 9 or 10 (see Figure 3, upper half). Conversely, higher grammar abilities predicted higher rates of accuracy in the metaphoric condition compared to the synonymic condition at 8, 9 and 10 years old, though not for the 6 and 7-year-olds (**Figure 5-3**, lower half). Interestingly, the only children in the entire sample to reach accuracy rates for the metaphoric-reference items that were comparable to those of the synonymic-reference items were those 10-year-olds who scored highly on grammar (see **Figure 5-3**, bottom right).

We note that the TOM index and language total score were themselves clearly correlated ( $r = .49, p < .001$ , see correlation



**Figure 5-4:** Predicted response times (in ms, on a logarithmic scale) as a function of age and condition, and partial residuals (residuals below 1000 ms or above 8000 ms [i.e., 1% of the data points] were not displayed and age was backtransformed from scaled values to improve readability).

matrix of all predictors in **Figure S2.1** in [supplementary materials](#)), but that the risk of multicollinearity for the model was limited (all *VIF* values < 2.33).

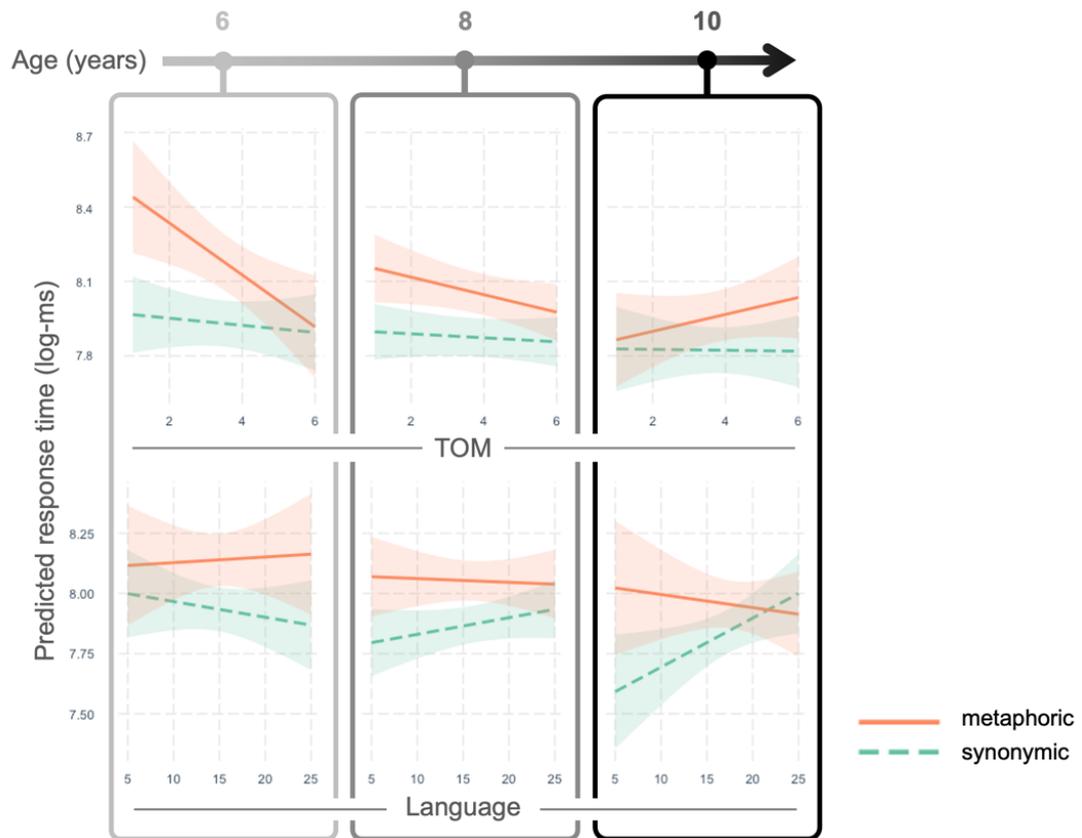
### Response times

Before running the RT analysis, we filtered the data set so that it included correct responses only (this amounts to removing 31% of the data points at this stage). We then removed outliers (4.6% of the remaining responses). The final set included 1880 log-transformed data points. In the first step, the quadratic model proved to provide a better fit to the data than the linear one ( $\chi^2(2) = 22.5, p < .001$ ), but adding a cubic term did not improve the fit ( $\chi^2(2) = 0.8, p = .70$ ). As can be seen in **Figure 5-4**, the selected quadratic model revealed a main effect of condition, with accurate post-metaphoric picture choices prompting longer response times compared to post-synonymic ones ( $\beta = 0.18, SE = 0.05, p < .01$ ). The estimated mean picture choice time for the post-synonymic control items was 2527 ms ( $SE = 105$ ) as opposed to 2879 ms ( $SE = 131$ ) for the metaphoric

ones. We will refer to the difference between these two conditions as the *metaphor-related cost*.

Reference condition was part of an interaction with age, so one can see that the metaphor-related cost evolved with age. Post-hoc tests revealed that the metaphor-related cost decreased between 6 and 7 years of age ( $\beta = 0.14, SE = 0.04, p < .001$ ) as well as between 7 and 8 ( $\beta = 0.09, SE = 0.02, p < .0001$ ) and marginally between 8 and 9 years of age ( $\beta = 0.04, SE = 0.02, p = .07$ ) but it did not change between 9 and 10 ( $p = .88$ ).

In a second step, we dropped the quadratic term for age and added the other covariates. The model's output is reported in **Table 5-3**. The model's estimates indicate that response times are involved in three-way interactions of TOM and grammar with age and condition. Higher TOM abilities are associated with faster responses after making a metaphoric reference as opposed to a synonymic one, but this influence decreased with age. The opposite pattern was observed for grammar, as for accuracy.



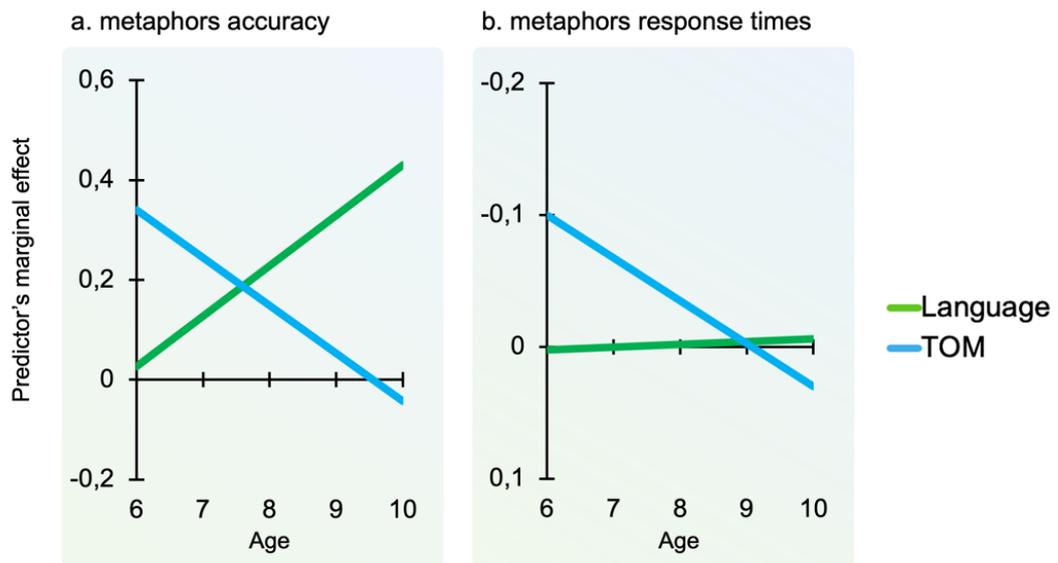
**Figure 5-5:** Predicted response times (in ms, log-transformed) across the two reference conditions, as a function of TOM (x-axis, top half) and grammar (x-axis, bottom half), at 3 different age points (6, 8 and 10 years old) (predictions were backtransformed from scaled variables).

Post-hoc contrasts (see **table S2.4** in [supplementary materials](#)) revealed that higher TOM predicted faster pictorial choice responses after making metaphoric rather than synonymic references at 6 and 7 years of age, but that no such effects were significant at older ages (**Figure 5-5**, upper half). Grammar was linked to response times of metaphoric as opposed to synonymic reference pictorial choices at 9 and 10, but not before. Visual inspection of the predictions (**Figure 5-5**, lower half) as well as estimates (**table S2.5** in [supplementary materials](#)) indicate that this effect is driven by a negative impact of language score on post-synonymic picture-choice response times at these ages (leading to longer response times), rather than being due to a positive impact of language on post-metaphoric picture-choice response times.

As far as how underlying factors affect the development of metaphor comprehension is concerned, **Figure 5-6** summarizes the marginal effects of TOM and language as a function of age on accuracy (left) and response times (right) in the metaphoric reference condition. One can see that TOM has a weakening influence on both accuracy and Picture-choice response times with age. Meanwhile, language competency has a facilitating effect on accurate metaphor picture choice with age.

### Exploratory analysis

To further explore the data and to characterize the nature of the post-reference picture choices, we fitted a supplementary model on the developmental trajectory model of response-times by adding – as a third response-type – responses that reflect the literal picture choice in the metaphor



**Figure 5-6:** Predicted partial effects of TOM and language on accuracy (left) and response times for correct responses (right) in the metaphoric reference condition, as a function of age (note that the y axis in (b) reverses polarity to indicate increased speed).

reference condition, which were previously considered as errors. In other words, we compare three kinds of response times: a) those that concern correctly choosing pictures among post-synonymic items, b) those that concern correctly choosing pictures in the wake of a metaphoric reference and; c) those that concern incorrectly choosing literal representations of the target reference pictures in the wake of the metaphorical-reference. This model showed that so-called *literal* readings of a metaphoric target reference were made more quickly than those that reflect correctly choosing pictures post-metaphorically, regardless of age (all  $ps < .001$ ) and. Moreover, these incorrect choices were made as fast as those found among correct post-synonymic items (all  $ps > .05$ , see **Table S2.6** and **Figure S2.3** in [supplementary materials](#)). This indicates that the choices reflecting literal interpretations in the metaphor reference condition were made precipitously and not ponderously.

## Discussion

In Experiment 2, we tested the tablet-based metaphor reference task on a larger group of children while presenting it as part of an assessment-like scenario. As anticipated, the results with respect to metaphor development here were consistent with those found in Experiment 1. Moreover, it provided finer detail since it included participants across the entire age range, i.e., between 6 and 10 years of age. We were thus able to show that metaphor comprehension development is non-linear and especially focused between 7 and 9 years old.

More germane to our investigation was determining the extent to which the development of TOM and grammar each appears to interact with metaphor development. Interestingly, the two cognitive abilities indeed appear to interact with development and in two different times in the chronological age of the participants. Theory of mind appears to have facilitative effects on making accurate metaphoric choices among the younger children. However, the effects of theory of

mind on metaphoric choices plateau by the time children are 9 years old. In contrast, grammar competence appears to have facilitative effects on the older children. Indeed, the only children to choose pictures correctly after a metaphoric reference at rates as high as the synonymic reference were the oldest children who scored highest for grammar.

The role of the two cognitive abilities with respect to reaction times was similarly revealing. For those children who accurately make metaphoric picture choices, we reported that the metaphor-related cost evolved with age. That is, the difference in metaphor-based and literal-based picture choice responding decreased between 6 and 9 years of age but it did not change between 9 and 10. Grammar was associated to response-time differences between metaphoric and synonymic references after 9 years of age. Much like in Experiment 1, this appears to be due to the fact that increased sophistication with respect to language impacts post-*synonymic* picture-choice response times in this age range (leading to longer response times in the control condition). That is, the older participants appear to slow down in the synonymic condition as they become aware of alternative meanings for the target reference rather than becoming faster due to a positive impact of language on post-metaphoric picture-choice response times.

Overall, one can see almost a hand-off between the two cognitive substrates. Early on, TOM is influential on picture choice in the metaphoric reference condition among the younger participants. This appears to wane by the time children are roughly 8 or 9 years old. At this point, grammar competence appears to take on an increasing influential role on children's metaphorical responding.

## General Discussion

We began this paper by providing background on prior metaphor development studies and by introducing a tablet-based referential metaphor task. In line with previous studies, data from Experiment 1 provided evidence indicating that metaphor comprehension develops with age and showed that metaphor-based picture selection, which reflects the study's dependent variable, took reliably longer than its controls. Importantly for the present paper, the task in Experiment 1 provided us with a proof of concept of the metaphor task's viability before it was juxtaposed with grammar and TOM tasks in a second experiment, that was much more comprehensive. That is, in Experiment 2, we included the metaphor-reference task as part of an assessment of pragmatic abilities that was distributed to a large sample of 238 typically-developing children, along with a non-verbal TOM task and a receptive grammar task. This second study allowed us to characterize the development of, and arguably the cognitive predictors of, referential metaphors. Not surprisingly, the data from the metaphor task in Experiment 2 was highly similar to those in Experiment 1. In the remainder of the General Discussion, we consider in greater detail what the task and the Experiments here bring to the rich literature on metaphor development. This detailed examination addresses the two general questions that motivated this research, one concerning the added value of the new variation of the task and another concerning what the adjacent cognitive tasks – on TOM and receptive grammar – reveal about their influence on metaphor comprehension development.

### **What does the current metaphor reference task provide for the literature?**

The developmental trend that appeared in both Experiments is consistent with a series

of results indicating that school-age is a key period for metaphor development. That is, all the children had little difficulty providing correct responses in the synonymic-reference control condition (which requires participants to choose the single picture that corresponds with the literal meaning of the target reference). This highlights how the linguistic material and the response format of our task is accessible to children. Meanwhile, picture choice-making in the metaphorical reference condition revealed greater equivocality. The data showed that incorrect picture choices in the metaphoric-reference condition typically reflects *literal* interpretations of the target reference. As children become older, they show improved metaphor comprehension performance, especially between the ages of 7 and 9 years, and without particular floor or ceiling effects. Up to about 8 to 9 years of age, the effort needed to choose metaphoric-based pictures as opposed to literally-based ones, in terms of response times, largely diminishes before stabilizing.

Although the developmental trend described in our Experiments is similar to those found in previous studies (Noveck et al., 2001; Seigneuric et al., 2016; Tonini et al., 2023), our rates of accuracy were somewhat lower. This can be explained by three methodological modifications made to the present study that are not found in the prior ones. First, we tested younger children while using oral, instead of written, materials. This has the advantage of avoiding interference with literacy because our youngest participants most likely are not accomplished readers. Second, we used a multiple-choice response format (instead of, say, questions requiring yes-no responses), which yields more variability. Importantly, this also allowed us to extend previous results by providing literal interpretations of metaphoric references as a picture choice option (see Van Herwegen

et al., 2013) and to positively identify a large proportion of participants' errors as non-random and literal. Third, we designed and selected metaphoric references that could then be made part of sensible initial contexts. Likewise, we built sensible contexts around our synonymic target references. In other words, by building up contexts around a sensible metaphoric or synonymic reference (rather than keeping a single context in which two different references are used), we were able to come up with highly reliable vignettes.

This study is also different from prior studies with respect to the way it records latency measures, when they exist. In prior studies, the dependent variable was the time needed to read a critical sentence that could contain a metaphoric or a literal reference. These perhaps directly grasp the effort associated with the online processing of the reference, but they could arguably be confounded with other reading-related procedures. In our task, we collect response times that occur downstream from the target reference. Interestingly, we still report metaphor related slowdowns. This shows that a) effects related to metaphor slowdowns are robust and that; b) it might not be related to referring back to a previously mentioned element. It could be the case that, for successful metaphor comprehenders, it is not reference-making in itself that is difficult but rather how to proceed once the metaphoric reference arrives. The current study appears to put the onus on what happens once a metaphorical reference arises. Future studies could further examine where the source of difficulty in a metaphoric reference task lies. Note that we would not necessarily hypothesize that metaphor-related slowdowns are merely costly; they are also argued to come with benefits by helping encoding and memorization (Noveck et al., 2001; Reynolds & Schwartz, 1983).

From the point of view of the current study's qualitative results, one finds consistency with both classical (Winner et al., 1976) and more recent studies (Deckert et al., 2019; Lecce et al., 2019). The current findings on age-related advances are also important because they are coherent with prior observations from other types of metaphor tasks. Our data are also in line with Willinger et al. (2019) and Deckert et al. (2019), who suggest that there are spurts in metaphor development among school aged children, as opposed to continuous development.

### **To what extent do theory of mind and language abilities provide scaffolding for metaphoric reference comprehension?**

We now turn to our second research question, which concerns identifying the potential role played by TOM and grammar in metaphor development. Before we begin, recall two features that served as the basis of our analyses. One is that we adopted a minimally-verbal measure of TOM (from Langdon & Coltheart, 1999) for the purposes of the current study. This is unlike TOM measures in other studies in this literature, which typically employ TOM assessments that are verbally rich. Consider the oft-used Strange Stories task (Happé, 1994), used in Lecce et al. (2019), which asks participants to detect subtle differences between jokes, white lies, misunderstandings and so on and through rather long vignettes and questions.<sup>20</sup> These sorts of tasks make it potentially difficult to disentangle the respective roles of language and theory of mind, which is central to our goal here (see de Villiers, 2007 on the interface between language and TOM). The

PST-based measure advances our understanding of the three-way relationship between metaphor comprehension, language and TOM abilities, as called for by Matthews et al. (2018). The other is that we used age as a continuous predictor in our statistical analysis, in interaction with the other variables, which allows us to exploit the whole sample's size and variability. This implies that the turning points we report should be considered as qualitative indications, rather than as exact values.

Our results reveal that TOM was positively associated with both accuracy and response times of picture choices after having heard metaphorical references as opposed to synonymic one. This TOM effect is clear at 6 years of age and steadily decreases with age, falling below significance threshold at roughly 9 years of age. This finding is in line with the hypothesis proposed in Lecce et al. (2019)'s study, which relied on tasks in which children metalinguistically explain a metaphor's meaning, and with Tonini et al. (2023) who, in addition, used a referential task. The current study shows that this observation can arguably be generalized because our data are generated by referential metaphors, tested with a multiple-choice format, and when TOM is assessed with our minimally verbal task. Hence, one can confidently conclude that the association we and others report between TOM and metaphor development is not linked to a language-based confound. Our paradigm also allowed us to test younger children and without reading, providing evidence that TOM's influence can be observed in children who are as young as 6-years-old. That said, the work here confirms and extends the hypothesis

---

<sup>20</sup> Likewise, consider the *Reading the Mind through the Eyes* (RME) task (Baron-Cohen et al., 2001), which asks participants to label mental states based on photographs of eyes;

assessments here rely on participants having a vocabulary that distinguishes between descriptions such as *serious*, *ashamed*, *alarmed*, *bewildered* and so on.

that says TOM has a passing influence on metaphor comprehension.

Formal language skills, in the form of receptive grammar, also revealed themselves to be statistically predictive of metaphorical reference behavior, even though this association was visible with respect to accuracy only (i.e., not to picture choice response times). This generally points to an important developmental shift during school age for the comprehension of metaphorical references: while young children capitalize on TOM, older children appear to be boosted by their receptive grammar skills. This pattern is consistent with Deckert et al. (2019) who observed (with another type of metaphor task) that verbal intelligence was associated with better metaphor performance in 9 and 10-year-olds, but not among those who are 7 or 8. They argued that “*early metaphors become gradually more linguistically articulated*” (p. 358).

What accounts for these two interlocking developmental trends? We consider two possibilities. The first is that children simply need to build strong TOM skills to understand metaphors (and potentially other phenomena) before eventually using those skills to later make refined linguistic judgements. Another possibility is that, within the grammar task, items that account for variations among younger children are not the same as those that distinguish abilities among older children. That is, simple items (for example, those targeting the distinction between past and present tenses) would discriminate young children but not older children, who are more likely to show refined abilities when comprehending complex structures (such as embedded clauses). Measures reflecting

complex grammatical competence, available to older children, would then avail themselves to tasks that measure metaphor comprehension. These two possibilities are not mutually exclusive.

The apparent developmental shift from TOM-supported metaphor comprehension among children to language-boosted comprehension has theoretical as well as applied implications. Regarding pragmatic theories in general, our results question a dichotomy recently formalized in the literature between linguistic pragmatics and social pragmatics (Andrés-Roqueta & Katsos, 2017, 2020). The first term designates pragmatic cases which depend on structural language and pragmatic norms, while the second relies on those plus competence with TOM. In this framework, scalar implicatures (cases in which weaker terms such as “some” prompt addressees to negate a more informative term such as “all” so that *some* is often pragmatically understood as *not all*) are typically viewed as a prime example of linguistic pragmatics while irony is considered exemplary of social pragmatics<sup>21</sup>. However, our results with metaphor show that a single pragmatic phenomenon (which these authors would classify as exemplary of linguistic-pragmatics) can be associated with both language and TOM, but at different times of development. While we agree with the authors quest to disentangle the umbrella term “pragmatic inferences” (for a large review see Noveck, 2018), our developmental results indicate that dividing the field up into phenomena is not necessarily the way to do it. At the very least, the current results underline the importance of considering developmental perspectives when building such models.

---

<sup>21</sup> The authors do specify that this dichotomy is not meant to apply to pragmatic phenomena *per se* but rather to communicative situations, and to

experimental paradigms, in that in some situation scalar implicature might rely on TOM abilities and vice-versa.

Despite our optimism that the work here advances discussion about metaphor development, we would be remiss if we did not consider the limitations of the current work. The first is that the present work has considered metaphors as a general phenomenon while not considering specific patterns that might emerge by distinguishing between its subtypes, such as the distinction between sensory as opposed to conceptual metaphors (Van Herwegen et al., 2013), or between physical and mental metaphors, where TOM is likely to have a greater importance for the second type (Lecce et al., 2019). Second, many prior studies used vocabulary measures as an index for language abilities, while we used grammar. This calls for caution when making comparisons among studies that underline linguistic abilities, and for further studies on the specific contribution of different language sub-components. Third, the second study relies on a cross-sectional correlational study; to be more convincing, the developmental trends we observe should be confirmed by longitudinal studies as well (see Del Sette et al., 2020). Finally, we observed *associations* between our variables of interest, rather than directly providing evidence for the causal nature of, and the direction of those relationships. Intervention studies, targeting either TOM of language to improve metaphor comprehension, should now help specifying the nature of those links, while also offering translational perspectives to clinical and educational settings.

To sum up, the current paper adds value to the rich literature on metaphor comprehension development among school-aged children in five critical ways. First, it provides yet another metaphor-reference task to the literature, whose outcomes – like its predecessors – show progressive improvement in metaphor comprehension with age. Second, the current task, unlike prior ones that largely

focus on the target reference itself, finely separates reference assignment from a picture selection portion of the task, showing that the metaphor comprehension and slowdown effects reported in the literature are robust, even as a participant's focus moves downstream (cf. Noveck et al., 2001). Thirdly, to our knowledge, this is the first study to consider the potential association of both structural language and TOM with metaphor comprehension with a single developmental sample. Specifically, the data show that, while young children capitalize on TOM, older children's improvement seems to take advantage of advancing skills at grammar. Fourthly, the TOM measure employed in the current paper is based on a non-verbal TOM task, making the TOM arguably less biased than those found in previous studies investigating its role in metaphor development. Finally, the task is carried out on a tablet that can be one of dozens carried into a classroom. This allows for congenial, and thus ecologically valid, data collection that engages young participants.

## References

- Almor, A., Arunachalam, S., & Strickland, B. (2007). When the Creampuff Beat the Boxer: Working Memory, Cost, and Function in Reading Metaphoric Reference. *Metaphor and Symbol*, 22(2), 169–193. <https://doi.org/10.1080/10926480701235478>
- 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, 996. <https://doi.org/10.3389/fpsyg.2017.00996>
- Andrés-Roqueta, C., & Katsos, N. (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](https://doi.org/10.1044/2020_JSLHR-19-00263)
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *J. Child Psychol. Psychiat.*, 42(2), 241–251.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/doi:10.18637/jss.v067.i01>.
- Bishop, D. V. M. (2003). Children’s Communication Checklist—2nd Edition. Pearson.
- Boyce, W., Torsheim, T., Currie, C., & Zambon, A. (2006). The Family Affluence Scale as a Measure of National Wealth: Validation of an Adolescent Self-Report Measure. *Social Indicators Research*, 78(3), 473–487. <https://doi.org/10.1007/s11205-005-1607-6>
- Carston, R., & Yan, X. (2022). Metaphorical Referring is Costly. *Book of Abstracts - XPrag 2022*. <https://doi.org/10.17605/OSF.IO/C4KP2>
- Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) Family Affluence Scale. *Social Science & Medicine*, 66(6), 1429–1436. <https://doi.org/10.1016/j.socscimed.2007.11.024>
- de Villiers, J. (2007). The interface of language and Theory of Mind. *Lingua*, 117(11), 1858–1878. <https://doi.org/10.1016/j.lingua.2006.11.006>
- Deckert, M., Schmoeger, M., Schaunig-Busch, I., & Willinger, U. (2019). Metaphor processing in middle childhood and at the transition to early adolescence: The role of chronological age, mental age, and verbal intelligence. *Journal of Child Language*, 46(2), 334–367. <https://doi.org/10.1017/S0305000918000491>
- Del Sette, P., Bambini, V., Bischetti, L., & Lecce, S. (2020). Longitudinal associations between theory of mind and metaphor understanding during middle childhood. *Cognitive Development*, 56, 100958. <https://doi.org/10.1016/j.cogdev.2020.100958>
- Di Sante, M., & Potvin, L. (2022). We Need to Talk About Social Inequalities in Language Development. *American Journal of Speech-Language Pathology*, 31(4), 1894–1897. [https://doi.org/10.1044/2022\\_AJSLP-21-00326](https://doi.org/10.1044/2022_AJSLP-21-00326)
- Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, 47–59.
- Gibbs, R. W. (1990). Comprehending figurative referential descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 56–66. <https://doi.org/10.1037/0278-7393.16.1.56>
- Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48(2), 101–119.
- Happé, F. (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. *Journal of Autism and Developmental Disorders*, 24(2), 129–154. <https://doi.org/10.1007/BF02172093>
- Heredia, R. R., & Cieśllicka, A. B. (Eds.). (2015). *Bilingual Figurative Language Processing*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139342100>
- Kalandadze, T., Bambini, V., & Næss, K.-A. B. (2019). A systematic review and meta-analysis of studies on metaphor comprehension in individuals with autism spectrum disorder: Do task properties matter? *Applied Psycholinguistics*, 40(6), 1421–1454. <https://doi.org/10.1017/S0142716419000328>
- Kalandadze, T., Norbury, C., Naerland, T., & Naess, K.-A. B. (2018). Figurative Language Comprehension in Individuals with Autism Spectrum Disorder: A Meta-Analytic Review. *Autism: The International Journal of Research and Practice*, 22(2), 99–117. <https://doi.org/10.1177/1362361316668652>

- Khomsi, A., Khomsi, F., & Pasquet, F. (2007). *BILO - Bilans Informatisés de Langage Oral*. Pearson.
- Kogan, N., & Chadrow, M. (1986). Children's comprehension of metaphor in the pictorial and verbal modality. *International Journal of Behavioral Development*, 9(3), 285–295.
- Kogan, N., Connor, K., Gross, A., & Fava, D. (1980). Understanding visual metaphor: Developmental and individual differences. *Monographs of the Society for Research in Child Development*, 1–78.
- Langdon, R., & Coltheart, M. (1999). Mentalising, schizotypy, and schizophrenia. *Cognition*, 71(1), 43–71. [https://doi.org/10.1016/S0010-0277\(99\)00018-9](https://doi.org/10.1016/S0010-0277(99)00018-9)
- Lecce, S., Ronchi, L., Del Sette, P., Bischetti, L., & Bambini, V. (2019). Interpreting physical and mental metaphors: Is Theory of Mind associated with pragmatics in middle childhood? *Journal of Child Language*, 46(2), 393–407. <https://doi.org/10.1017/S030500091800048X>
- Lenth, R. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (R package version 1.8.2). <https://CRAN.R-project.org/package=emmeans>
- Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual Differences in Children's Pragmatic Ability: A Review of Associations with Formal Language, Social Cognition, and Executive Functions. *Language Learning and Development*, 14(3), 186–223. <https://doi.org/10.1080/15475441.2018.1455584>
- Morsanyi, K., Stamenković, D., & Holyoak, K. J. (2020). Metaphor processing in autism: A systematic review and meta-analysis. *Developmental Review*, 57, 100925. <https://doi.org/10.1016/j.dr.2020.100925>
- 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>
- Noveck, I. (2018). *Experimental Pragmatics: The Making of a Cognitive Science*. Cambridge University Press. <https://doi.org/10.1017/9781316027073>
- Noveck, I., Bianco, M., & Castry, A. (2001). The Costs and Benefits of Metaphor. *Metaphor and Symbol*, 16(1–2), 109–121. <https://doi.org/10.1080/10926488.2001.9678889>
- Piaget, J. (1964). *Le langage et la pensée*. Éditions Gonthier SA.
- Porter, M. A., Coltheart, M., & Langdon, R. (2008). Theory of Mind in Williams Syndrome Assessed Using a Nonverbal Task. *Journal of Autism and Developmental Disorders*, 38(5), 806–814. <https://doi.org/10.1007/s10803-007-0447-4>
- Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics*, 25, 51–79. <https://doi.org/10.1075/bjl.25.04pou>
- Pouscoulous, N., & Tomasello, M. (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, 156, 160–167. <https://doi.org/10.1016/j.pragma.2019.05.021>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rajkumar, A. P., Yovan, S., Raveendran, A. L., & Russell, P. S. S. (2008). Can only intelligent children do mind reading: The relationship between intelligence and theory of mind in 8 to 11 years old. *Behavioral and Brain Functions*, 4(1), 51. <https://doi.org/10.1186/1744-9081-4-51>
- Reynolds, R. E., & Schwartz, R. M. (1983). Relation of metaphoric processing to comprehension and memory. *Journal of Educational Psychology*, 75, 450–459. <https://doi.org/10.1037/0022-0663.75.3.450>
- Rosseel, Y. (2012). **lavaan**: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>
- Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in*

- Medicine*, 25(1), 127–141. <https://doi.org/10.1002/sim.2331>
- Seigneuric, A., Megherbi, H., Bueno, S., Lebahar, J., & Bianco, M. (2016). Children's comprehension skill and the understanding of nominal metaphors. *Journal of Experimental Child Psychology*, 150, 346–363. <https://doi.org/10.1016/j.jecp.2016.06.008>
- Sperber, D., & Wilson, D. (1996). *Relevance: Communication and Cognition* (2nd ed.). Wiley-Blackwell.
- Tonini, E., Bischetti, L., Del Sette, P., Tosi, E., Lecce, S., & Bambini, V. (2023). On the specificity of the relationship between metaphor skills and Theory of Mind in middle childhood: Reviving the Referential Metaphors task. *Cognition*.
- Van Herwegen, J., Dimitriou, D., & Rundblad, G. (2013). Development of novel metaphor and metonymy comprehension in typically developing children and Williams syndrome. *Research in Developmental Disabilities*, 34(4), 1300–1311. <https://doi.org/10.1016/j.ridd.2013.01.017>
- Van Rheenen, T. E., & Rossell, S. L. (2013). Picture sequencing task performance indicates theory of mind deficit in bipolar disorder. *Journal of Affective Disorders*, 151(3), 1132–1134. <https://doi.org/10.1016/j.jad.2013.07.009>
- Vézina, M., Sylvestre, A., & Fossard, M. (2013). Development of a Quebec French Version of the. *Revue canadienne d'orthophonie et d'audiologie*, 37(2), 156–168.
- Vosniadou, S. (1987). Children and metaphors. *Child Development*, 870–885.
- Whyte, E. M., & Nelson, K. E. (2015). Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *Journal of Communication Disorders*, 54, 2–14. <https://doi.org/10.1016/j.jcomdis.2015.01.001>
- Willinger, U., Deckert, M., Schmöger, M., Schaunig-Busch, I., Formann, A. K., & Auff, E. (2019). Developmental steps in metaphorical language abilities: The influence of age, gender, cognitive flexibility, information processing speed, and analogical reasoning. *Language and Speech*, 62(2), 207–228.
- Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, 12, 289–297. <https://doi.org/10.1037/0012-1649.12.4.289>

## Point d'étape à l'issue de l'étude 2

Après avoir étudié le développement des implicatures scalaires à l'âge scolaire dans une *première étude*, cet article se focalise sur le développement de la compréhension des métaphores dans cette fenêtre développementale, au cours de laquelle la littérature décrit des développements tardifs, malgré des preuves attestant la présence de précurseurs précoces avant cela, à l'âge préscolaire. De plus, deux prédicteurs principaux de la compréhension de métaphores ont notamment émergé des travaux sur l'autisme : la théorie de l'esprit et les compétences langagières formels. Toutefois, ces deux prédicteurs ont été relativement peu étudiés dans le développement typique, et surtout pas de façon conjointe, et avec des tâches expérimentales de théorie de l'esprit mobilisant fortement les compétences verbales, et induisant ainsi un risque de biais important. Cette étude répond à ces limitations en introduisant une nouvelle tâche clinique, dans le cadre du TIPi. Cette tâche développée sur tablette évalue la compréhension de références métaphoriques, par opposition à des références synonymiques, à l'aide de matériel oral, et permet la collecte des temps de réponse. Dans une première expérience, nous avons confirmé la validité de surface et la sensibilité développementale de cette nouvelle tâche auprès de 89 enfants de 7 et 10 ans. Dans une seconde expérience, nous avons utilisé une version standard de cette tâche à 248 enfants à développement typique entre 6 et 11 ans, ainsi qu'une tâche validée de compréhension de phrase, et une tâche non-verbale de théorie de l'esprit minimalement verbale, que nous avons validée pour une utilisation clinique pédiatrique dans une *étude annexe*. Les résultats ont révélé un développement non-linéaire de la compréhension des métaphores, avec une amélioration des performances focalisée entre 7 et 9 ans. Par ailleurs, des modèles mixtes expliquant les réponses et les temps de réponses des enfants ont mis en évidence une bascule développementale dans les prédicteurs cognitifs de la compréhension de métaphores. En effet, la théorie de l'esprit constituait un prédicteur significatif des performances en métaphores chez les enfants les plus jeunes, mais cette association diminuait avec l'âge, laissant la place à une association grandissante avec les compétences langagières formelles, qui constituaient un prédicteur fiable chez les enfants les plus âgés. Cette étude souligne donc l'importance d'adopter une perspective développementale dans la considération des prédicteurs cognitifs des habiletés métaphoriques, et des modèles pragmatiques en général. Dans le cadre de cette thèse, elle ouvre également sur un certain nombre de considérations cliniques qui pourront être développées ultérieurement. Conjointement avec la *première étude*, ces données constituent en outre des points de repères développementaux, à la lumière desquels peuvent être observés des développements atypiques, et en particulier autistiques. C'est la démarche que nous adopterons dans la *prochain chapitre* de cette thèse, dans lequel nous présenterons des résultats cliniques préliminaires issu de la confrontation de ces données aux performances d'un échantillon d'enfants autistes.

## Chapitre 6 • Données cliniques préliminaires (étude 3)

Dans les chapitres précédents, des repères développementaux ont été constitués en évaluant la compréhension des implicatures scalaires puis des métaphores chez un échantillon relativement large d'enfants à développement typique. Ces repères peuvent désormais être mobilisés afin d'étudier le développement autistique des inférences pragmatiques, au sujet duquel un certain nombre de questions émergent de la littérature. Ce chapitre présentera ainsi des données cliniques préliminaires, dans le prolongement des réflexions et méthodologies de cette thèse. Après avoir rappelé brièvement les questionnements qui nous motivent, nous présenterons la méthodologie adoptée ici, puis les résultats obtenus, avant de discuter de leurs implications. Des informations plus détaillées sont présentées dans un manuscrit d'article en préparation, disponible en [annexe D](#).

### Introduction

Dans le [chapitre 3](#), nous avons vu qu'en dépit des représentations caricaturales associées à l'autisme, évoquant un déficit pragmatique global, la littérature expérimentale apparaît plus nuancée. Un contraste semble notamment se dessiner entre des phénomènes pragmatiques qui seraient atypiques, notamment les métaphores (Morsanyi et al., 2020), et d'autres typiques, comme les implicatures scalaires (e.g. Chevallier et al., 2010). Toutefois, ce contraste a peu été observé au sein d'un même échantillon de personnes autistes, et chacune de ces lignes de recherche présente des limitations qui appellent à la prudence.

Pour ce qui est des métaphores, l'observation de différences entre les performances de participants autistes et neurotypiques a été répliquée dans de nombreuses études (Kalandadze et al., 2018, 2019; Morsanyi et al., 2020), mais les tailles d'effet rapportées varient largement. Ces variations ont notamment été expliquées par les stratégies d'appariement des groupes et par les tâches expérimentales utilisées (Kalandadze et al., 2018; Morsanyi et al., 2020), qui feraient varier l'influence du facteur 'langage' dans ces effets. Des tâches d'explications verbales libres donnent classiquement lieu à des effets de taille plus importante (Kalandadze et al., 2018), mais sont aussi critiquées pour leur importante demande métalinguistique (Pouscoulous, 2011). Certains auteurs proposent donc que les difficultés classiquement associées à l'autisme soient mieux expliquées par des difficultés verbales plutôt que par des caractéristiques proprement autistiques, et seraient ainsi également présentes chez les enfants avec un trouble du langage (Norbury, 2005).

Pour ce qui est des implicatures scalaires, une majorité d'études a échoué à mettre en évidence des différences entre participants autistes et neurotypiques, adultes ou enfants

(Andrés-Roqueta & Katsos, 2020; Chevallier et al., 2010; Hochstein et al., 2017; Pijnacker et al., 2009; Schaeken et al., 2018; Su & Su, 2015). Toutefois, ces études ont utilisé des tâches de jugement binaire qui, comme nous l'avons discuté dans les chapitres précédents, ne peuvent pas être considérées comme une mesure directe de la compréhension d'implicatures (Katsos & Bishop, 2011; Katsos, 2014; Kissine & De Brabanter, 2023). Deux études ont utilisé d'autres modalités d'évaluation et tendent à mettre en évidence des différences (Mazzaggio et al., 2021; Pastor-Cerezuela et al., 2018), mais des limitations méthodologiques empêchent encore de tirer des conclusions claires (voir la section du [chapitre 3](#) dédiée à cette discussion).

Ainsi, le contraste entre ces deux types d'inférences apparaît plus trouble qu'initialement. Ceci fragilise également des réflexions théoriques qui s'appuient sur cette distinction, comme la dichotomie entre pragmatique linguistique et pragmatique sociale (Andrés-Roqueta & Katsos, 2017, 2020). Par ailleurs, hormis quelques exceptions, ce contraste a plutôt été relevé en comparant les résultats de différentes études, plutôt qu'observé au sein d'un même échantillon. Ainsi, ce chapitre propose d'exploiter les tâches expérimentales introduites dans les chapitres précédents afin d'évaluer (i) si des enfants autistes se comportent de façon typique face à des énoncés sous-informatifs comparables à ceux utilisés dans les tâches de jugement binaires, mais dans une tâche expérimentale évaluant *directement* ces implicatures scalaires, avec un mode de réponse non-verbal ; (ii) si des enfants autistes réalisent des performances plus faibles qu'attendu dans une tâche de compréhension de métaphores qui limite et contrôle la demande verbale et métalinguistique ; et enfin (iii) si le contraste entre implicatures scalaires typiques et métaphores atypiques peut être observé au sein d'un même échantillon d'enfants autistes.

## Méthodologie

### Population

Vingt-trois enfants autistes (3 filles) ont été recrutées au Centre Hospitalier Le Vinatier pour cette étude, âgés de 6;7 ans à 10;8 ans. Chacun avait reçu un diagnostic de Trouble du Spectre de l'Autisme sans trouble du développement intellectuel ni trouble du langage par un psychiatre de l'enfant et de l'adolescent et remplissait les critères diagnostiques du DSM-5. Ce diagnostic était confirmé par des évaluations diagnostiques de références (ADOS-2, SCQ et/ou ADI-R) extraites des dossiers médicaux des enfants. Des évaluations récentes du fonctionnement intellectuel n'étaient pas toujours disponibles (données manquantes pour 3 enfants), et ont été réalisées avec des outils variés, mais celles-ci révèlent toutefois chez cet échantillon un QI total dans la zone normale à normale-supérieure (moyenne = 111, IC à 95% = [100 ; 122], données manquantes = 10), avec une intelligence verbale dans la zone normale-supérieure (moyenne = 107, IC à 95% = [101 ; 113], données manquantes = 4), tout comme l'intelligence fluide (moyenne = 113, IC à 95% = [106 ; 120], données

manquantes = 4). Un trouble de déficit de l'attention / hyperactivité (TDAH) associé était considéré comme critère d'exclusion. Tous les enfants étaient régulièrement scolarisés et avaient le Français comme langue maternelle (le bilinguisme n'était pas disqualifiant).

Afin de constituer un groupe contrôle neurotypique, les données de 237 enfants des études présentées dans les *chapitres 4 et 5 (expérience 2)* ont été extraites. Pour rappel, cet échantillon a été recruté dans 2 écoles primaires. Les parents de ces enfants ne rapportaient pas de trouble du neurodéveloppement. Par ailleurs, cet échantillon n'incluait que des enfants dont les performances en langage formel et en théorie de l'esprit ne déviaient pas de plus de 2,5 écart-types sous la moyenne de leur groupe d'âge (voir détails dans l'*expérience 2 du chapitre 5*).

Tous les participants et leurs parents ont fourni un consentement éclairé à participer à cette étude, qui a reçu un avis favorable du Comité de Protection des Personnes Sud-Est I (ID RCB 2019-A01721-56).

### **Matériel**

Les habiletés langagières formelles des enfants ont été évaluées avec deux tests français validés issues de la batterie BILO-3C (Khomsi et al., 2007), afin de confirmer l'exclusion de troubles du langage, et de mieux caractériser le profil des participants autistes. Nous avons plus particulièrement évalué leur compréhension lexicale à l'aide d'une épreuve de jugement lexical produisant un score brut total variant de 0 à 43, et leur compréhension grammaticale à l'aide d'une épreuve d'appariement phrase-image fournissant un score brut total variant de 0 à 27.

Afin d'évaluer leurs inférences pragmatiques, les enfants ont été testés avec la tâche d'implicature scalaire présentée au *chapitre 4* et la tâche de références métaphoriques présentée au *chapitre 5, expérience 2*.

### **Procédure**

Les participants autistes ont été évalués dans un bureau calme de l'hôpital, individuellement. Les différentes tâches ont été proposées dans un ordre personnalisé, afin de s'adapter à la motivation et à la disponibilité de chaque enfant, et en plusieurs séances si nécessaire (généralement 2 visites). Les enfants neurotypiques, comme présenté dans les chapitres 4 et 5, ont été évalués collectivement dans leur classe. Pour ce groupe, la tâche de métaphores, plus longue, était proposée avant la tâche d'implicatures scalaires. Au sein de chaque tâche, les items étaient proposés selon des séquences standardisées, identiques pour chaque enfant de cette étude.

### **Analyses**

Une approche par trajectoire développementale (Thomas et al., 2009) a été utilisée pour l'analyse des résultats. Cette approche consiste à régresser les performances par l'âge dans chaque groupe, et à comparer les trajectoires qui en résultent. L'avantage

de cette méthode est qu'elle fournit des informations plus riches qu'une simple comparaison de moyennes, et permet de tirer parti de l'hétérogénéité d'un groupe (notamment en termes d'âge), plutôt que d'être limité par cette dispersion. Nous l'avons dans un premier temps appliquée aux mesures contrôles de langage formel, avec des régressions linéaires multiples, puis aux tâches expérimentales avec des modèles linéaires mixtes. Des modèles mixtes généralisés ont été utilisés pour évaluer les réponses (correctes vs incorrectes). Pour ce qui est des temps de réponse, une approche plus classique de comparaison de moyennes a été adoptée, compte-tenu de la forte variabilité associée à ce type de variables dépendantes. Pour ce faire, des modèles linéaires mixtes ont été ajustés, après élimination des réponses incorrectes, transformation logarithmique pour corriger l'asymétrie des distributions, et exclusion des valeurs extrêmes déviant de plus de 2,5 écart-types de la moyenne par groupe et par condition. Chaque set d'analyse suit les logiques développées distinctement pour les implicatures scalaires dans le [chapitre 4](#) et pour les métaphores dans le [chapitre 5, expérience 2](#).

## Résultats

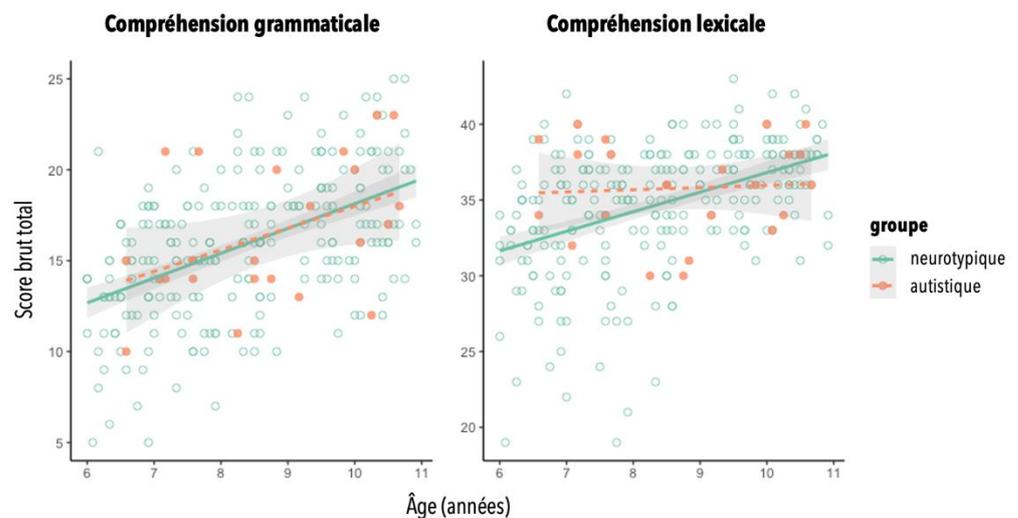
### Mesures contrôles de langage

Le tableau 6-1 présente les mesures démographiques et la description des performances des deux groupes aux différentes tâches.

**Tableau 6-1** : Caractéristiques de l'échantillon et descriptions des performances de chaque groupe pour les tâches contrôles et expérimentales

		<i>Neurotypique</i>	<i>Autistique</i>	Différence inter-groupes
	N	237	23	
Informations démographiques	Filles	51%	13%	$X^2(1, N = 260) = 0.00, p < .05$
	Âge moyen (ET)	8.4 (1.4)	8.7 (1.4)	$t(258) = 0.98, p = .33$
	SES Moyen (ET)	58.8 (15.7)	53.2 (17.0)	$t(258) = 1.62, p = .11$
Tâches contrôle de langage formel	Vocabulaire réceptif moyen (ET)	34.7 (4.2)	35.8 (3.2)	
	Grammaire réceptive Moyenne (ET)	15.9 (3.8)	16.5 (3.9)	
Métaphores	Performance moyenne, condition littérale (ET)	95% (12%)	94% (11%)	
	Performance moyenne, condition métaphores (ET)	44% (35%)	49% (37%)	
Implicatures scalaires	Taux d'erreurs sémantiques moyen (ET)	2% (0.6%)	2% (0.5%)	
	Taux de réponses pragmatiques moyen (ET)	79% (34%)	65% (43%)	

Pour ce qui est du langage formel, les trajectoires développementales des deux groupes en compréhension grammaticale sont remarquablement superposées (figure 6-1, gauche), les analyses montrant un effet principal clair de l'âge ( $\beta = 1.36$ ,  $SE = 0.16$ ,  $p < .001$ ), mais pas du groupe ( $\beta = 1.8$ ,  $SE = 4.8$ ,  $p = .70$ ) et pas d'interaction entre les deux facteurs ( $\beta = -0.20$ ,  $SE = 0.54$ ,  $p = .71$ ). En compréhension lexicale (figure 6-1, droite), on observe un effet principal de l'âge ( $\beta = 1.3$ ,  $SE = 0.17$ ,  $p < .001$ ) et du groupe ( $\beta = 10.8$ ,  $SE = 5.35$ ,  $p < .05$ ) avec une interaction approchant la significativité ( $\beta = -1.15$ ,  $SE = 0.6$ ,  $p < .10$ ), les enfants autistes tendant à obtenir de meilleures performances que les enfants neurotypiques dans la section la moins âgée de l'échantillon.



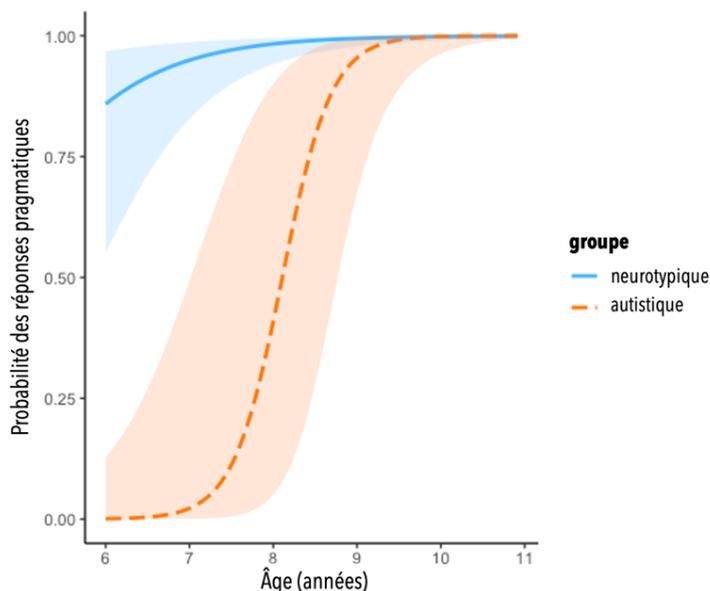
**Figure 6-1 :** performances individuelles et régressions linéaires des participants aux épreuves de compréhension grammaticale (gauche) et lexicale (droite) en fonction de l'âge et du groupe ; IC à 95% représentés par les zones grisées

### Implicatures scalaires

Pour ce qui est des performances à la tâche d'implicatures scalaires, nous avons dans un premier temps considéré les erreurs sémantiques, qui était rares (2% des réponses). Le modèle statistique montre par ailleurs qu'elles sont légèrement moins fréquentes avec l'âge ( $OR = 0.67$ ,  $95\% CI = [0.48, 0.93]$ ,  $p < .05$ ), mais qu'il n'y a pas d'effet du groupe ( $OR = 0.58$ ,  $95\% CI = [0.15-2.22]$ ,  $p > .05$ ), ni d'interaction entre les deux facteurs ( $OR = 2.39$ ,  $95\% CI = [0.60-9.48]$ ,  $p > .05$ ). Ces résultats confirmant que la tâche est aussi accessible aux enfants autistes qu'aux enfants neurotypiques, nous pouvons nous tourner vers la condition d'intérêt : les réponses aux énoncés sous-informatifs.

Dans cette condition critique, les réponses pragmatiques étaient variables (77% des réponses au total). Le modèle statistique a par ailleurs révélé que la probabilité des réponses pragmatiques était influencée par l'âge ( $OR = 4.87$ ,  $95\% CI = [2.4, 9.9]$ ) mais aussi par le groupe ( $OR = 0.03$ ,  $95\% CI = [0.00, 0.38]$ ,  $p < .01$ ), avec une interaction significative entre le groupe et l'âge ( $OR = 24.7$ ,  $95\% CI = [2.0, 297.7]$ ,  $p < .05$ ). Comme le révèle la figure 6-2, les réponses pragmatiques étaient en effet clairement

moins fréquentes dans le groupe d'enfants autistes, et cette différence était particulièrement marquée chez les enfants les plus jeunes.



**Figure 6-2 :** Probabilité des réponses pragmatiques prédites par le modèle statistique en fonction de l'âge et du groupe

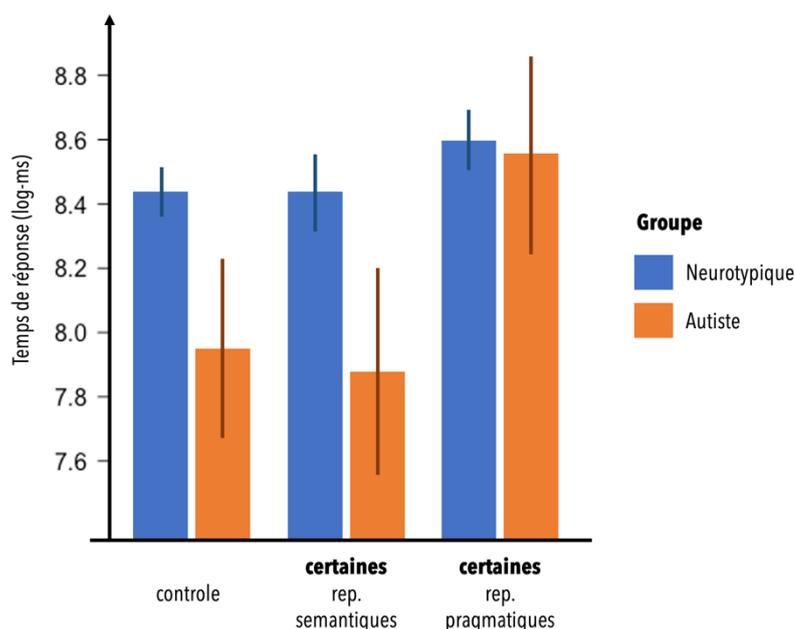
Pour ce qui est des temps de réponse, nous nous intéressons à l'effet du type de réponse (réponses pragmatiques à la condition critique / réponses sémantiques à la condition critique / réponse aux quantifieurs contrôles) indépendamment des variables contrôles (voir le [chapitre 4](#) pour une description plus détaillée de cette méthodologie). Cette analyse a révélé une interaction entre le type de réponse et le groupe (voir les paramètres du modèles exacts dans le tableau 6-2).

**Tableau 6-2 :** Paramètres du modèle expliquant les temps de réponse (après transformation logarithmique) dans la tâche d'implicatures scalaires

Prédicteurs	B	CI À 95%	P
(Intercept)	8.02	7.95 - 8.09	<0.001
hasMoved	0.06	-0.10 - 0.22	0.435
<b>moves</b>	<b>0.22</b>	<b>0.20 - 0.25</b>	<b>&lt;0.001</b>
RESPONSE-TYPE1	-0.17	-0.37 - 0.03	0.102
RESPONSE -TYPE2	-0.16	-0.40 - 0.07	0.173
<b>groupe [autistes]</b>	<b>-0.37</b>	<b>-0.57 - -0.16</b>	<b>&lt;0.001</b>
RESPONSE -TYPE1 * groupe [autistes]	-0.36	-1.03 - 0.31	0.296
<b>RESPONSE -TYPE2 * groupe [autistes]</b>	<b>-0.58</b>	<b>-1.00 - -0.16</b>	<b>0.007</b>

Les analyses post-hoc révèlent le même pattern dans les deux groupes (voir figure 6-3) : les réponses pragmatiques sont plus lentes que les réponses sémantiques à la condition critique et que les réponses aux quantifieurs contrôles (tous les  $ps < .05$ ), mais cette différence était amplifiée chez les participants autistes. Ainsi, les participants autistes étaient plus rapides que les participants neurotypiques en cas de

réponse sémantique à la condition critique ( $\beta = 0.56$ ,  $SE = 0.14$ ,  $p < .001$ ) ou de réponse aux quantifieurs contrôles ( $\beta = 0.48$ ,  $SE = 0.13$ ,  $p < .001$ ), mais ils étaient aussi lents que les contrôles en cas de réponse pragmatique ( $\beta = 0.05$ ,  $SE = 0.14$ ,  $p > .05$ ).



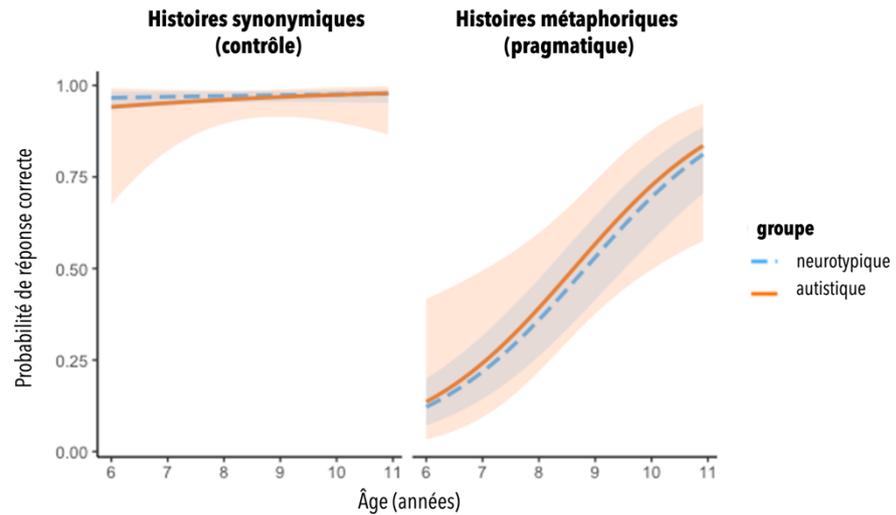
**Figure 6-3** : Effets partiels du groupe et du type de réponse sur les temps de réponse (après transformation logarithmique) d'après le modèle statistique ; les barres d'erreurs représentent les intervalles de confiance à 95%

## Métaphores

Dans la tâche de métaphores, les taux de réponses correctes étaient très hauts dans la condition contrôle synonymique (95%) et plus variés dans la condition critique métaphorique (45%), où la plupart des erreurs résultaient de la sélection des propositions de réponse littérales (94% des erreurs). Le modèle statistique (tableau 6-3) révèle une interaction très claire entre l'âge et la condition mais pas d'effet du groupe, ni d'interaction du groupe avec les autres facteurs. En d'autres termes, les trajectoires développementales des deux groupes étaient particulièrement superposées dans les deux conditions (voir figure 6-4).

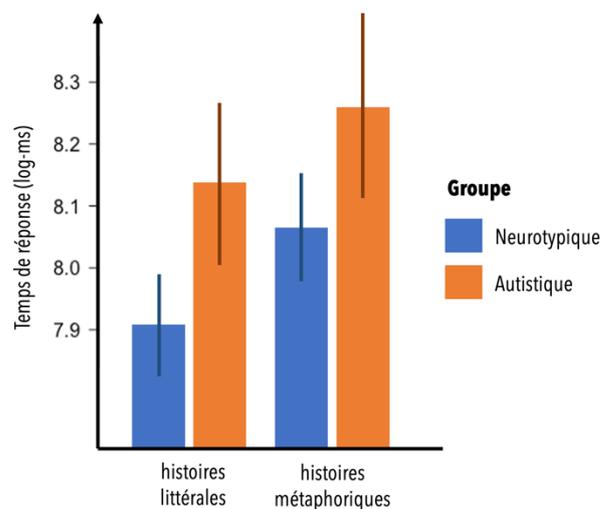
**Tableau 6-3** : Paramètre du modèle statistique mixte généralisé expliquant les réponses correctes dans la tâche de métaphore

Prédicteurs	Odds Ratios	CI à 95%	p
(INTERCEPT)	5.01	3.48 - 7.20	<0.001
<b>CONDITION</b>	<b>46.75</b>	<b>24.14 - 90.52</b>	<b>&lt;0.001</b>
<b>AGE</b>	<b>1.76</b>	<b>1.43 - 2.18</b>	<b>&lt;0.001</b>
GROUPE	0.87	0.43 - 1.75	0.691
<b>CONDITION * AGE</b>	<b>0.43</b>	<b>0.32 - 0.58</b>	<b>&lt;0.001</b>
CONDITION * GROUPE	0.70	0.28 - 1.76	0.443
AGE * GROUPE	1.21	0.57 - 2.56	0.622
(CONDITION * AGE) * GROUPE	1.18	0.42 - 3.29	0.754



**Figure 6-4 :** Probabilité de réponse correcte prédite par le modèle généralisé en fonction du groupe, de l'âge et de la condition

Pour ce qui est des temps de réponse, le modèle statistique a été ajusté sur un jeu de données de 2125 points après l'exclusion de 2% de valeurs extrêmes. Il révèle un effet principal de la condition ( $\beta = -1.16$ ,  $SE = 0.06$ ,  $p < .05$ ) et du groupe ( $\beta = 0.21$ ,  $SE = 0.05$ ,  $p < .001$ ), mais pas d'interaction entre les facteurs ( $\beta = 0.03$ ,  $SE = 0.06$ ,  $p > .05$ ). Comme le révèle la figure 6-5, la condition métaphore était associée à des temps de réponse plus longs, et les participants autistes étaient plus lents de manière générale. L'absence d'interaction révèle que le coût temporel associé à la condition métaphorique par rapport à la condition contrôle était comparable entre les deux groupes.



**Figure 6-5 :** Temps de réponse (après transformation logarithmique) prédits par le modèle en fonction du groupe et de la condition; les barres d'erreurs représentent les intervalles de confiance à 95%

## Discussion

Dans ce chapitre, nous avons comparé les données de développement typique rapportées dans les chapitres 4 et 5 aux performances d'enfants autistes dans de nouvelles tâches d'implicatures scalaires et de compréhension de métaphores. Nous cherchions à évaluer si, comme ils le montrent avec des tâches de jugement binaire, des enfants autistes se comportent de façon typique face à des implicatures scalaires, dans une tâche écologique qui ne fait pas appel à des processus métalinguistiques explicites. Nous voulions également déterminer si ces enfants présentaient des difficultés pour comprendre des références métaphoriques dans un paradigme limitant et contrôlant la charge langagière et métalinguistiques qu'il induit. Ceci s'inscrivait dans une volonté plus large de rechercher un contraste entre ces différents types d'inférences pragmatiques. Cependant, à l'inverse de nos hypothèses, nous avons observé chez les enfants autistes de moins bonnes performances que chez leurs pairs neurotypiques dans la tâche d'implicatures scalaires, mais des performances similaires en compréhension de métaphores. L'observation des réponses comme des temps de réponse étaient cohérents avec ce point de vue, puisque les enfants autistes ont montré un coût plus important des réponses pragmatiques par rapport aux situations contrôle dans la tâche scalaire, mais un coût temporel similaire associé aux métaphores. Dans la suite de ce chapitre, nous détaillerons d'abord comment ces résultats peuvent s'inscrire dans la littérature sur les implicatures scalaires d'une part et sur les métaphores d'autre part, puis nous examinerons pourquoi ce contraste était inattendu ; nous considérons enfin la nécessité d'une explication alternative à nos observations.

### Implicatures scalaires

Nos résultats contrastent donc avec la plupart des études précédentes, qui avaient observé que des participants autistes traitaient les implicatures scalaires de façon typique, enfants comme adultes (Andrés-Roqueta & Katsos, 2020; Chevallier et al., 2010; Hochstein et al., 2017; Pijnacker et al., 2009; Schaecken et al., 2018; Su & Su, 2015), quand nos données suggèrent que les implicatures scalaires sont moindrement dérivées par les enfants autistes, et à un coût plus important. Les participants autistes de notre échantillon ne semblaient pourtant pas en difficulté avec la tâche elle-même. De plus, les caractéristiques de cet échantillon rendent peu probable que ces différences soient expliquées par des particularités autres que le diagnostic d'autisme, puisque les enfants de l'échantillon clinique présentaient de bonnes compétences verbales et intellectuelles. Ce groupe différait seulement du groupe contrôle quant à sa distribution en genre, mais les données rapportées au [chapitre 4](#) montraient que ce facteur n'avait pas d'incidence sur les performances dans cette tâche. Une différence importante entre notre paradigme et les tâches de jugement binaire classiques est cependant qu'il demande une utilisation relativement *spontanée* de l'informativité pour inférer l'intention de communication du locuteur, comme développé dans le [chapitre 4](#). Des participants autistes jugent autant que des participants neurotypiques un énoncé sous-informatif comme faux (Andrés-Roqueta & Katsos,

2020; Chevallier et al., 2010; Hochstein et al., 2017; Pijnacker et al., 2009; Schaeken et al., 2018; Su & Su, 2015), mais dans notre situation où de tels énoncés sont présentés sans consignes spécifiques, sous forme de demandes, ils sont plus nombreux à considérer qu'une interprétation dite sémantique pourrait correspondre aux intentions du locuteur. Une façon d'intégrer nos résultats aux travaux précédents serait donc de considérer que les personnes autistes seraient typiquement capables d'évaluer si un énoncé sous-informatif est pragmatiquement approprié lorsqu'on le leur demande (bien que de manière atypique lorsqu'on leur donne la possibilité d'un jugement intermédiaire, voir Schaeken et al. [2018]), mais qu'elles seraient moins susceptibles d'utiliser *spontanément* l'informativité pour dériver les intentions des locuteurs, ce qui entraînerait moins d'interprétations pragmatiques.

Les temps de réponse de nos participants sont également en faveur de cette hypothèse. En effet, les études précédentes ne mettaient pas non plus en évidence de différences inter-groupes sur ce type de variable dépendante. À l'inverse, nos données montrent que quand bien même les enfants autistes réalisaient des interprétations pragmatiques, cela était manifestement opéré au prix d'un coût temporel plus important que les enfants neurotypiques, par rapport à leur propre vitesse de réponse aux items contrôles.

Nos résultats soutiennent tout du moins cette explication dans un contexte développemental, ce qui pourrait ou non évoluer à l'adolescence ou à l'âge adulte. Cette hypothèse est par ailleurs renforcée par l'observation de taux de réponses pragmatiques qui tendent à être réduits chez les enfants autistes lorsqu'ils sont évalués avec des paradigmes de questions à choix multiple (Mazzaggio et al., 2021 ; Pastor-Cerezuela et al., 2018). Nous examinerons plus tard comment cette possible explication pourrait être intégrée à une hypothèse plus large sur la compréhension pragmatique dans l'autisme.

## **Métaphores**

Nos données vont aussi à l'encontre de précédents travaux qui mettaient en évidence un accès entravé – ou du moins atypique – aux métaphores dans l'autisme (Kalandadze et al., 2018, 2019; Morsanyi et al., 2020), alors que nous observons des trajectoires développementales très similaires entre nos groupes. Il nous paraît peu probable que l'effectif relativement modeste de notre échantillon ait échoué à révéler une différence pourtant bien présente, pour trois raisons. La première est que la plupart des études antérieures mobilisaient des échantillons d'effectifs comparables (voir par exemples les études incluses dans la méta-analyse de Morsanyi et al. en 2020). La deuxième est que la tâche que nous avons employée ici montrait pour autant une très bonne sensibilité développementale chez des enfants à développement typique (voir [chapitre 5](#) ou figure 6-4). Enfin, ce même échantillon se distinguait clairement des participants neurotypiques dans une autre tâche pragmatique, celle évaluant les implicatures scalaires, comme discuté ci-dessus. La littérature expérimentale antérieure est toutefois caractérisée par une grande variabilité dans la taille des effets qu'elle rapporte, ce qui a été mis en lien avec les stratégies d'appariement avec les groupes

contrôles (Kalandadze et al., 2018) et à des variables expérimentales (Kalandadze et al., 2019; Morsanyi et al., 2020): les effets étaient typiquement plus faibles quand les groupes étaient appariés en niveau de langage et quand les paradigmes utilisaient des réponses par choix multiples plutôt que des explications verbales libres. Notre étude remplit ces deux critères, ce qui rend ses résultats moins surprenants, même si un effet était tout de même attendu. Un possible facteur explicatif supplémentaire est que les enfants autistes que nous avons évalués présentaient un léger avantage en termes d'intelligence non-verbale sur le groupe contrôle, ce qui aurait pu dissimuler des différences inter-groupes subtiles. Cependant, des données récentes ont montré que le lien entre compréhension des métaphores et raisonnement analogique chez des participants autistes disparaissait dès lors que les habiletés langagières (vocabulaire, ici) étaient prises en compte (Morsanyi et al., 2022). Une autre hypothèse explicative est que la tâche que nous avons utilisée limite la demande linguistique, non seulement dans la modalité de réponse, mais également de tout le matériel utilisé, avec du vocabulaire fréquent et une faible complexité syntaxique, comme le confirment les performances au plafond dans la condition contrôle. Ces hypothèses ne s'excluent pas mutuellement, mais quoi qu'il soit nos résultats confirment que des participants autistes peuvent interpréter des métaphores de façon typiques, dans ces conditions.

Les temps de réponse peuvent par ailleurs être utilisés pour mieux comprendre les processus qui sous-tendent cette compréhension. Quelques études précédentes avaient exploité ce type de mesure auprès de participants autistes, et révélaient des différences inter-groupes, même dans des cas où les réponses en elles-mêmes étaient comparables (Chahboun et al., 2016; Morsanyi et al., 2022). Cependant, si ces études rapportaient des temps de traitement d'un matériel métaphorique plus longs dans l'autisme que chez des participants neurotypiques, elles n'incluaient pas de matériel contrôle impliquant seulement une compréhension littérale. Ces paradigmes ne permettaient donc pas de distinguer si les participants autistes étaient plus lents pour traiter les métaphores, ou pour réaliser la tâche en général, ce que notre paradigme est en mesure de faire. Nos données révèlent ainsi que les participants autistes étaient plus lents que les enfants neurotypiques dans les items métaphoriques, mais également dans les items contrôles, sans interaction entre la condition et le groupe. En d'autres termes, le coût temporel associé aux métaphores par rapport aux histoires contrôles synonymiques, décrit au [chapitre précédent](#), était similaire dans les deux groupes. Cela suggère que des processus atypiques étaient recrutés par la tâche de façon très générale, plutôt que par les métaphores spécifiquement, explication qui pourrait également s'appliquer aux résultats des études précédentes. En cohérence, les quelques études précédentes qui incluaient des conditions contrôles ne mettaient pas en évidence d'interaction claire entre le groupe et la condition (Gold et al., 2010; Gold & Faust, 2010).

### **Un contraste au sein des inférences pragmatiques**

Notre question de recherche plus générale était de déterminer si un contraste entre la compréhension des implicatures scalaires et des métaphores pouvait être observé au sein d'un même échantillon d'enfants autistes. Sur la base de la dichotomie entre

linguistique pragmatique et linguistique sociale, qui propose que les performances de participants autistes dépendent de la demande en théorie de l'esprit de la situation, nous attendions que les participants autistes divergent des enfants neurotypiques sur la tâche de pragmatique sociale (métaphores) mais pas de pragmatique linguistique (implicatures scalaires). Pourtant, comme discuté ci-dessus, le contraste que nous avons observé était opposé à celui attendu. Face à ces résultats, il faut souligner dans un premier temps que, comme décrit par Marocchini (2023), ce n'est pas parce qu'une tâche est réussie par des participants autistes qu'elle n'implique pas de raisonnement social, ou plus spécifiquement de théorie de l'esprit, en dépit de l'héritage de la théorie de la cécité mentale. En effet, nous avons montré au [chapitre 5](#) que la tâche de métaphore que nous avons introduite était associée à la théorie de l'esprit (notamment chez les enfants les plus jeunes), et elle s'est pourtant révélée bien réussie par les participants autistes que nous avons évalués. Il nous semble donc que le contraste que nous avons mis en évidence appelle une autre dichotomie.

Comme discuté dans la littérature sur les métaphores, diminuer les demandes verbales et interactionnelles d'une situation est typiquement considéré comme un facteur facilitateur chez les participants autistes (Kalandadze et al., 2019; Morsanyi et al., 2020). C'est pourtant ce que nous avons proposé dans chacune tâches expérimentales, et pas uniquement dans la tâche de métaphore, et cela ne peut donc être utilisé pour expliquer le contraste que nous avons observé. En revanche, un autre facteur nous semble distinguer ces deux tâches : la tâche d'implicatures scalaires nécessite une utilisation *spontanée* de l'information pour inférer les intentions de communication du locuteur, tandis que la tâche de métaphores propose des questions et des propositions de réponses qui peuvent *guider* ou structurer le raisonnement pragmatique. Sur la base de cette distinction, nous interrogeons donc la possibilité que les participants autistes montrent des performances pragmatiques typiques dans un paradigme qui guide et structure ce raisonnement, mais qu'ils rencontrent plus de difficultés pour utiliser ces compétences spontanément ; ils réaliseraient alors des performances moindres dans des tâches expérimentales qui guident moins ce raisonnement, et rencontreraient des difficultés dans leurs interactions au quotidien (avec des locuteurs principalement neurotypiques). Il nous semble que cette distinction pourrait trouver une place au sein des littératures sur les métaphores et sur les implicatures scalaires, en aidant à y intégrer nos résultats.

Elle pourrait notamment contribuer à expliquer que les tâches de métaphores utilisant des explications verbales libres soient celles qui produisent les effets les plus importants. Jusqu'alors, seule la demande verbale de ce format de réponse était utilisée pour justifier l'importance de ces effets. Par exemple, la demande verbale a été utilisée pour expliquer que l'étude de Rundblad et Annaz (2010) rapporte l'effet le plus large de la méta-analyse de Kalandadze et al. (2019). Nous constatons toutefois avec intérêt que cette étude employait une tâche relativement similaire à celle de cette thèse, puisqu'elle utilisait des histoires verbales se terminant par une métaphore, suivie par une question sur le référent de cette métaphore. Comme dans notre tâche, ces questions

portaient sur des éléments concrets et évitaient des termes mentaux (par exemple « Que voit Stuart ? » après qu'une foule ait été décrite comme une « marée »). Une différence importante entre nos paradigmes, toutefois, est que nous avons affiché des propositions de réponse, tandis que Rundblad et Annaz demandaient aux enfants de répondre oralement à la question, sans propositions. Bien que Kalandadze et al. (2019) aient catégorisé cette tâche d'« *explication verbale libre* » (p. 1441), les enfants n'étaient pas tenus ici d'*expliquer* cette métaphore, et pouvaient se contenter de répondre d'un seul mot. La demande verbale de cette tâche n'était donc pas excessivement plus importante que dans la nôtre, mais pourtant à l'origine de résultats très différents. Nous proposons donc qu'une question ouverte guide beaucoup moins le raisonnement pragmatique que des propositions de réponse, et que ce facteur pourrait jouer un rôle important dans cette différence entre nos résultats.

Par ailleurs, d'autres travaux ont précédemment montré que des modifications subtiles dans les paradigmes expérimentaux pouvaient avoir des impacts importants sur les différences inter-groupes d'études sur l'autisme. A.C. Wilson & Bishop (2021) ont par exemple montré que des adultes autistes à qui l'on présentait des questions sur des significations implicites étaient bien plus enclins à sélectionner une proposition de réponse « Je ne sais pas » que des adultes non-autistes. Toutefois, lorsque cette possibilité était retirée, forçant les participants à sélectionner une interprétation parmi plusieurs, ils répondaient le plus souvent de façon similaire aux participants contrôles. Ce résultat suggère donc que le raisonnement pragmatique des participants autistes serait lié à une préférence générale pour la certitude, mais aussi que des propositions de réponse pourraient *guider* ce raisonnement, contraignant les participants autistes à atteindre une conclusion, alors qu'ils pourraient spontanément s'arrêter sur un résultat incertain.

Inversement, en comparaison des tâches de jugement classique, le paradigme basé sur l'action que nous avons utilisé pour évaluer les implicatures scalaires diminue la demande métalinguistique. Ce faisant, il améliore les performances d'enfants à développement typique (Pouscoulous et al., 2007 ; *chapitre 4*). Mais alors que cette manipulation devrait aussi rendre la tâche plus accessible aux enfants autistes, elle a produit l'effet inverse : leurs performances étaient relativement entravées par cette manipulation. Ceci indique que la demande métalinguistique ne peut pas rendre compte de nos résultats, à la différence de l'hypothèse que nous proposons.

Bien sûr, cette hypothèse ne constitue qu'une explication a posteriori des résultats que nous avons obtenus, hypothèse qui nécessiterait désormais d'être spécifiquement testée. Une manière de la mettre à l'épreuve serait de comparer les performances d'individus autistes dans des tâches spontanées / implicites vs guidées / explicites d'un même phénomène pragmatique, tout en contrôlant la demande verbale et métalinguistique des tâches. Plus généralement, ces données cliniques préliminaires présentent des limitations auxquelles de futures recherches devraient répondre. Premièrement, il est possible que les patterns que nous avons observés évoluent avec

l'âge, mais que la taille réduite de notre échantillon n'ait pas permis de l'observer. De plus, un échantillon plus large pourrait révéler des effets plus subtils, tels que des patterns développementaux non-linéaires, comme dans la *deuxième étude* de cette thèse, ou des effets plus fins sur les temps de réponse (notamment sur les métaphores), ce type de données étant typiquement plus bruité que des réponses simples. De plus, nous avons pris soin de sélectionner des participants autistes de façon à contrôler un certain nombre de variables confondantes, en excluant plusieurs comorbidités, mais dans la vie réelle l'autisme est le plus souvent associé à d'autres conditions (Levy et al., 2010), et la généralisabilité de nos observations à d'autres sous-groupes d'individus autistes doit donc être évaluée.

### Conclusion et point d'étape à l'issue de l'étude 3

Nous avons vu dans le *chapitre 3* que l'autisme était classiquement associé à des difficultés dans les inférences pragmatiques, ce qui a principalement été objectivé avec le langage figuré, et notamment les métaphores. Toutefois, l'étude plus récente des implicatures scalaires a majoritairement échoué à mettre en évidence des différences autistiques, mettant en doute l'hypothèse de difficultés pragmatiques générales dans l'autisme. De plus, chaque ligne de recherche présente des limitations : d'une part, la demande langagière et métalinguistique des tâches de métaphores pourrait expliquer des différences inter-groupes, plutôt que des différences dans la compréhension des métaphores en elles-mêmes ; d'autre part, les implicatures scalaires ont principalement été évaluées avec des tâches de jugement binaires, qui auraient pu biaiser les performances des participants.

Cette étude exploite donc les tâches expérimentales introduites dans les chapitres précédents (*chapitre 4* pour les implicatures scalaires, *chapitre 5* pour les métaphores) et qui répondent à ces limitations, afin de déterminer si un tel contraste peut alors toujours être observé, au sein d'un même échantillon clinique. Vingt-trois enfants autistes d'âge scolaire, sans trouble du langage, ni trouble du développement intellectuel, ni trouble du déficit de l'attention ont donc été recrutés et testés, et leurs performances ont été comparées à celles de 237 enfants neurotypiques des études précédentes. De façon surprenante, en comparaison de ce groupe contrôle, notre échantillon clinique était caractérisé par un contraste *inverse* à celui attendu. Plus précisément, les enfants autistes ont montré une compréhension plus pauvre et plus coûteuse des implicatures scalaires que les enfants neurotypiques, mais des performances comparables dans la tâche de références métaphoriques.

En complément de facteurs explicatifs précédemment introduits dans la littérature, nous interrogeons la possibilité que ce contraste soit expliqué par une opposition entre des processus pragmatiques *guidés* et *spontanés*. Parallèlement à des hypothèses formulées pour la théorie de l'esprit, les personnes autistes pourraient montrer des

capacités pragmatiques typiques, lorsque mesurées à l'aide de paradigmes *guidants*, mais pourraient rencontrer des difficultés à recruter ces compétences *spontanément*. Cela se traduirait par des performances moindres dans les tâches qui ne guident pas le raisonnement pragmatique et par des difficultés dans les interactions de la vie quotidienne, qui exigent en outre de monitorer les états mentaux d'autrui dans un environnement bien plus bruyant.

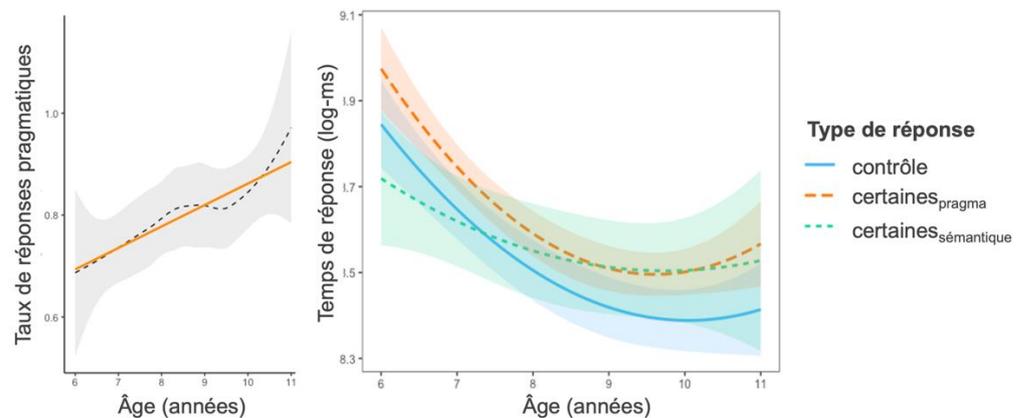
Dans le *prochain chapitre* de discussion générale, nous développerons l'intégration de ces résultats et de ceux des chapitres précédents dans la littérature scientifique, ainsi que leurs implications théoriques comme pratiques.

# DISCUSSION GÉNÉRALE

# Chapitre 7 • Synthèse et discussion des résultats obtenus

## 1. Synthèse des résultats

Dans cette thèse, nous avons proposé deux nouvelles tâches expérimentales au sein d'un nouvel outil d'évaluation clinique, le TIPi, qui visent à évaluer la compréhension des implicatures scalaires et des métaphores chez l'enfant d'âge scolaire. Chacune de ces tâches limite la charge langagière qu'elle induit, et la contrôle en incluant à la fois des items expérimentaux impliquant les phénomènes d'intérêt, et des items contrôles appariés mais ne nécessitant qu'une compréhension littérale. Par ailleurs, les formats de réponse de chaque tâche limitent la demande métalinguistique, avec pour la tâche d'implicatures scalaires une réponse par l'action, et pour la tâche de références métaphoriques des questions à choix multiples d'images. Enfin, chaque tâche est automatisée sur tablette tactile et permet le recueil des temps de réponse des enfants, favorisant aussi leur motivation et leur engagement. Ces deux tâches ont été appliquées auprès d'une cohorte relativement large d'enfants à développement typique (*chapitre 4* et *chapitre 5*) et d'un échantillon d'enfants autistes (*chapitre 6*).



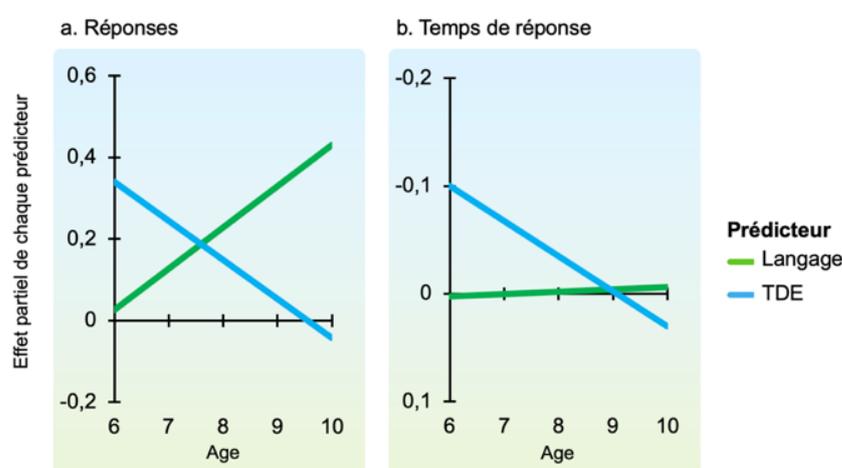
**Figure 7-1:** Principaux résultats de l'étude 1, représentant l'évolution avec l'âge des taux de réponses pragmatiques individuels des enfants tout-venant (gauche) et de leurs temps de réponse en fonction du type de réponse (droite)

La nouvelle tâche d'implicatures scalaires que nous avons proposée, adaptée sur tablette de celle de Pouscoulous et al. (2007), permet d'évaluer directement comment des termes scalaires sont interprétés par des enfants, à la différence des tâches de jugement binaire classiquement utilisées. En effet, dans cette tâche les participants sont confrontés à des demandes incluant des termes scalaires, telles que « *Je veux que certaines assiettes aient une fraise* » ; leurs réponses concrètes à ces requêtes révèlent donc plus directement les intentions de communication qu'ils attribuent au locuteur, et donc leur interprétation de 'certaines'. La *première étude* de cette thèse, auprès de

249 enfants à développement typique, a confirmé que cette tâche était largement accessible même aux enfants les plus jeunes, avec des taux d'erreurs sémantiques très bas et indépendants de l'âge. En revanche, les items sous-informatifs critiques suscitaient plus de variabilité, avec un taux de réponse pragmatique global de 79%, relativement haut, et une probabilité de répondre pragmatiquement qui augmentait linéairement avec l'âge entre 6 et 11 ans (figure 7-1, gauche). Enfin, les temps de réponses montraient que les réponses pragmatiques à la condition critique étaient associées à des latences, en comparaison des réponses aux items contrôles, et des réponses sémantiques (au moins chez les enfants les plus jeunes, à 6-7 ans, voir figure 7-1, droite). Cette différence semblait se réduire avec l'âge, et chez les enfants les plus âgés les réponses sémantiques tendaient à devenir également plus lentes que les réponses contrôles, bien que cette différence ne soit que marginalement significative. Ce faisant, nous avons confirmé le développement continu des implicatures scalaires au cours de l'âge scolaire, et montré que celui-ci témoigne plus probablement, au moins chez les enfants les plus jeunes, d'une immaturité dans la mise en place des inférences pragmatiques que d'une tolérance pragmatique vis-à-vis d'énoncés suboptimaux.

La tâche de références métaphoriques, inspirée de celles de Noveck et al. (2001) et de Van Herwegen et al. (2013), proposait des histoires verbales orales à la fin desquelles une expression était utilisée pour faire référence à un élément introduit plus tôt, référence qui pouvait être métaphorique ou synonymique selon le contexte initial. Une question permettait ensuite de capturer la compréhension de cette référence par les participants, grâce à plusieurs propositions de réponse sous forme d'images, incluant notamment la représentation d'une interprétation métaphorique et d'une interprétation littérale. Cette nouvelle tâche a été validée dans une première expérience de l'*étude 2* auprès de 89 enfants à développement typique, montrant qu'une référence synonymique ne posait pas de difficulté à des enfants mêmes jeunes, tandis qu'il était plus difficile pour les participants d'identifier le référent de ces mêmes expressions lorsque le contexte les rendait métaphoriques ; des erreurs plus fréquentes et des réponses plus lentes étaient alors observées. Les enfants de 10 ans montraient toutefois des difficultés nettement moindres que des enfants de 7 ans, confirmant la sensibilité développementale de la tâche. Dans une seconde expérience, un sous-ensemble standard de ces items a été proposé à un échantillon plus large de 238 enfants à développement typique, afin d'évaluer si la variation interindividuelle en compréhension des métaphores était associée aux variations interindividuelles en langage formel (compréhension grammaticale) et en théorie de l'esprit. La théorie de l'esprit a notamment été évaluée à l'aide d'une tâche minimalement verbale, sur tablette également, validée dans une étude annexe (*annexe B*) et répondant à un manque de la littérature. Les résultats montraient qu'à la différence des implicatures scalaires, les performances des enfants ne progressaient pas de façon linéaire avec l'âge, avec une amélioration marquée entre 7 et 9 ans, mais moindre avant et après ces âges.

Les modèles statistiques expliquant la variance dans les performances et dans les temps de réponses à cette tâche montraient des interactions entre l'âge, la condition (métaphore ou contrôle) et les prédicteurs cognitifs d'intérêt. Dans la condition métaphorique par rapport à la condition contrôle, l'association avec la théorie de l'esprit diminuait avec l'âge, tandis que l'association avec le langage formel augmentait. En d'autres termes, une bonne compréhension des métaphores était associée à une bonne théorie de l'esprit dès 6 ans et jusqu'à environ 8 ans, mais plus après, tandis qu'à l'inverse elle était associée à un bon langage formel seulement à partir de 8 ans environ, et jusqu'à 10 ans. Ces effets étaient visibles dans les réponses des enfants, ainsi que, dans une moindre mesure, dans leurs temps de réponse (figure 7-2). Ces résultats confirment donc l'implication à la fois des habiletés langagières et de la théorie de l'esprit dans la compréhension des métaphores, telle qu'évaluée par cette tâche. Plus précisément, ils confirment l'influence décroissante de la théorie de l'esprit avec l'âge (Lecce et al., 2019; Tonini et al., 2023) et montrent que ce phénomène est à l'inverse associé à une influence croissante des compétences langagières formelles. Ces éléments constituent ainsi une *basculée développementale* pendant l'âge scolaire, suggérant un changement qualitatif et pas seulement quantitatif dans les mécanismes de traitement des métaphores.

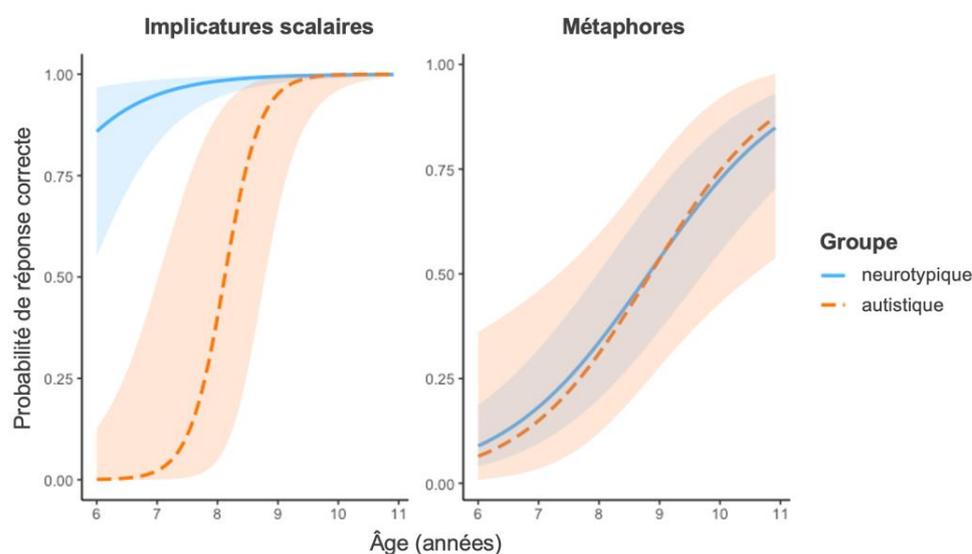


**Figure 7-2 :** Effets partiels des prédicteurs cognitifs (langage et théorie de l'esprit) sur les performances (à gauche) et les temps de réponse (à droite) des enfants dans la condition test de la tâche de métaphores, en fonction de l'âge (étude 2, expérience 2)

Considérées conjointement, ces deux premières études confirment donc que le développement des inférences pragmatiques est encore actif à l'âge scolaire, malgré les précurseurs précoces qui peuvent être observés tant pour la compréhension des implicatures scalaires (e.g., Katsos & Bishop, 2011) que des métaphores (e.g., Pouscoulous & Tomasello, 2020), sans que ces résultats ne puissent être attribués à la charge langagière ou métalinguistique des tâches, ici limitée et contrôlée.

Finalement, ces données obtenues auprès d'enfants à développement typique ont pu être confrontées aux performances d'un échantillon de 23 enfants autistes évalué avec ces deux mêmes tâches, dans une *troisième étude*. Ces enfants présentaient un autisme

relativement isolé, sans difficultés langagières ni intellectuelles, et sans déficit attentionnel. Ce dispositif visait à évaluer i) si la préservation des implicatures scalaires dans l'autisme était confirmée par une tâche relativement écologique, sans jugement métalinguistique 'artificiel', ii) si les difficultés en compréhension de métaphores étaient maintenues lorsque la demande métalinguistique et langagière de la tâche était faible, et iii) si le contraste entre implicatures scalaires préservées et métaphores entravées pouvait être observé au sein d'un même échantillon. Contrairement à nos hypothèses, les résultats de cette étude ont mis en évidence que les enfants autistes présentaient des performances moindres qu'attendu dans la tâche d'implicatures scalaires (figure 7-3, gauche), associées à un coût temporel des réponses pragmatiques par rapport aux items contrôles qui était majoré par rapport au groupe neurotypique. À l'inverse, la trajectoire développementale de ce groupe dans la tâche de métaphores était tout à fait similaire à celle observée dans le développement typique (figure 7-3, droite). Les temps de réponses des enfants dans cette tâche révélaient un ralentissement général chez les enfants autistes, mais sur l'ensemble des items, y compris les histoires contrôles impliquant une compréhension seulement littérale. L'absence d'interaction avec la condition (métaphore ou contrôle) ne permettait pas de conclure à un quelconque rapport atypique aux métaphores dans notre échantillon. Finalement, un contraste entre implicatures scalaires et métaphores a donc bien été observé, mais dans le sens inverse à celui attendu, les enfants autistes manifestant des performances plus faibles dans la tâche d'implicatures scalaires, mais pas dans la tâche de métaphores.



**Figure 7-3 :** Trajectoire développementales des réponses des groupes neurotypique et autistique dans les conditions tests des tâches d'implicatures scalaires (à gauche) et de métaphores (à droite), issue de l'étude 3

Chacun de ces résultats ayant été discuté isolément dans les chapitres qui leur étaient consacrés, nous proposons dans la suite de ce chapitre d'examiner plus globalement leurs implications et les perspectives qu'ils ouvrent quant aux modèles pragmatiques dans un premier temps, en revenant sur les notions de spontanéité des processus

pragmatiques, et sur la variabilité de leurs socles cognitifs. Dans un second temps, nous discuterons de l'apport possible de ces données à notre compréhension de l'autisme, et des pistes qu'elles ouvrent pour la pratique clinique.

## 2. Implications pour les modèles pragmatiques

### 2.1. Inférences pragmatiques spontanées ou guidées

La *première étude* de cette thèse était largement motivée par l'hypothèse de la tolérance pragmatique (Katsos & Bishop, 2011) et la critique méthodologique qu'elle porte à l'encontre des tâches de jugement binaires classiquement utilisées dans la littérature sur les implicatures scalaires (*chapitre 2*). Si ces tâches montrent un effet développemental clair (e.g. Barner et al., 2011; Katsos et al., 2016; Noveck, 2001; Papafragou & Musolino, 2003), Katsos et Bishop opposent en effet que ces mesures ne peuvent pas être strictement utilisées comme des mesure directes des implicatures scalaires (voir aussi Kissine & De Brabanter, 2023), puisque d'autres stratégies peuvent conduire à accepter ou rejeter un énoncé sous-informatif. Katsos et Bishop ont alors introduit une tâche de jugement ternaire, et démontré que dès 5 ans, les enfants peuvent se montrer aussi sensibles que des adultes à l'informativité d'un énoncé, ce qui est un prérequis important pour la compréhension des implicatures scalaires. Bien qu'il n'ait pas été développé dans ce but, le paradigme de Pouscoulous et al. (2007), que nous avons adapté sur tablette tactile, permet d'aller plus loin en mesurant cette fois *directement* la compréhension de l'implicature scalaire. Par exemple, face à une table et plusieurs fraises, un enfant à qui l'on demande « *Je veux que certaines assiettes aient une fraise* » ne peut se contenter de détecter la sous-informativité de cet énoncé compte-tenu du contexte, ou de se focaliser sur sa vérité propositionnelle, mais est contraint d'inférer les intentions du locuteur pour répondre efficacement à sa demande. Cette tâche s'approche donc de situations naturelles de la vie quotidienne. L'effet développemental que nous avons observé confirme que les enfants les plus jeunes comprennent moins l'implicature scalaire véhiculée par cet énoncé que les plus âgés, répliquant à une large échelle les résultats obtenus avec des tâches de jugement binaire. Il paraît peu probable que les enfants (au moins les plus jeunes) aient détecté la sous-informativité des énoncés en cas de réponses sémantiques, puisque ces réponses étaient plus rapides qu'en cas d'interprétations pragmatiques, dans lesquelles la sous-informativité doit être détectée et utilisée pour inférer l'intention de communication du locuteur.

Nous envisageons une façon de concilier ces résultats apparemment contradictoires, en proposant que les enfants puissent juger l'informativité d'un énoncé *lorsqu'on leur demande explicitement*, mais qu'ils échouent à utiliser *spontanément* cette information pour inférer l'intention de communication d'un locuteur dans des tâches plus implicites, telles que la nôtre. Ainsi, l'effet développemental classiquement observé pourrait s'expliquer (en partie) par un défaut de sollicitation ou d'attention spontanée

à l'informativité d'un énoncé, plutôt qu'une incapacité propre à percevoir ou traiter cette information.

Comme argumenté par Pouscoulous et al. (2007), la nature métalinguistique des tâches classiques de jugement binaire contribue également probablement aux difficultés que rencontrent les enfants qui s'y prêtent. Ainsi, les taux de réponses pragmatiques sont plus élevés dans notre tâche dont le mode de réponse est entièrement non-verbal. Cependant, les résultats de notre *troisième étude* renforcent l'hypothèse que ce facteur ne soit pas le seul qui distingue notre tâche des tâches classiques. En effet, si notre tâche semble faciliter les implicatures scalaires chez les enfants à développement typique, elle semble au contraire augmenter les difficultés (relatives) que rencontrent les enfants autistes. Il apparaît peu probable que limiter la demande d'une tâche augmente sa difficulté dans cet échantillon, d'autant que la charge métalinguistique d'une tâche est considérée comme une barrière particulière dans l'autisme (*chapitre 3*). Ces observations poussent donc à rechercher un autre facteur distinctif. La spontanéité avec laquelle les informations et processus doivent être mobilisés pour réaliser l'inférence pragmatique apparaît comme un candidat plausible pour distinguer ces tâches. Selon cette hypothèse, les opérations pragmatiques mobiliseraient des prérequis cognitifs, nécessaires mais non suffisants, qui devraient être recrutés de façon opportune, dans la vie de tous les jours. Ce recrutement serait alors plus ou moins spontané ou guidé, selon les situations, et notamment selon les paradigmes expérimentaux qui sont déployés en laboratoire. Cette hypothèse fait notamment écho à la littérature sur la théorie de l'esprit, qui a de la même façon contrasté des évaluations implicites et explicites des fausses croyances (Senju, 2012), mobilisant une compréhension guidée ou spontanée. Par rapport aux évaluations explicites, les évaluations implicites se sont similairement montrées accessibles aux enfants plus jeunes (Southgate et al., 2007), mais étaient aussi à l'origine de difficultés spécifiques chez des participants autistes (Senju et al., 2009). Cette hypothèse formulée a posteriori pourrait désormais être directement évaluée en comparant les performances de participants dans différentes tâches évaluant le même phénomène pragmatique, tel que les implicatures scalaires, en faisant varier le caractère implicite ou explicite de la tâche (manipulant la spontanéité des processus), éventuellement de façon orthogonale à leur charge métalinguistique. Nous développerons plus amplement l'utilité potentielle de ce concept pour les modèles de l'autisme dans la suite de chapitre.

## **2.2. Des bases cognitives multiples**

Outre l'étude de la trajectoire développementale de la résolution de références métaphoriques, la *deuxième étude* présentée permettait d'évaluer l'importance de deux prédicteurs cognitifs potentiels de cette compréhension : la compréhension grammaticale et la théorie de l'esprit. Cette étude permettait ainsi de répondre aux limitations des études précédentes, qui ne se focalisaient que sur l'un ou l'autre de ces prédicteurs, et utilisaient des tâches de théorie de l'esprit fortement verbales, limitant leur capacité à isoler la contribution réelle de la théorie de l'esprit, indépendamment

du langage. Cette question prenait une importance particulière au regard de la dichotomie entre pragmatique linguistique (mobilisant les compétences langagières) et pragmatique sociale (mobilisant la théorie de l'esprit), issue notamment des travaux sur l'autisme (*chapitre 3*). Andrés-Roqueta et Katsos (2017), décrivent cette opposition comme s'appliquant à des situations pragmatiques, plutôt qu'à des phénomènes en tant que tels, mais nos résultats poussent à nuancer plus encore cette proposition. En effet, une unique situation expérimentale est ici associée à des prédicteurs différents *selon les individus*, et plus particulièrement selon leur âge : la théorie de l'esprit prédit les performances des enfants les plus jeunes, et s'efface au profit du langage formel chez les enfants les plus âgés. Formulé dans les termes de cette dichotomie, avec une même situation, cette tâche évaluerait la pragmatique sociale chez les enfants les plus jeunes, et la pragmatique linguistique chez les enfants les plus âgés, ce qui paraît peu probable. De plus, cette dichotomie est mobilisée pour expliquer les performances des participants autistes, mais les prédicteurs semblent décorrélés de la présence ou non de difficultés chez les participants autistes de l'*étude 3*. On n'observe ainsi pas plus de difficultés avec les métaphores chez les enfants autistes plus jeunes (à la différence des implicatures scalaires, *figure 7-3*), à un âge où la théorie de l'esprit est pourtant le prédicteur le plus important dans la *deuxième étude*. Si cette dichotomie capture donc une certaine variabilité observée dans la littérature, tant dans les prédicteurs des performances pragmatiques que dans les résultats de participants autistes, nos résultats poussent donc à la nuancer et à dépasser ses limites.

Tandis que les raisons de cette bascule développementale ont pu être discutées dans le *chapitre 5* consacré à cette étude, cette observation suggère plus globalement que différentes stratégies ou bases cognitives puissent être mobilisées pour réaliser une même tâche pragmatique. Appliqué au modèle de la théorie de la pertinence, qui spécifie plusieurs sources d'information pour le module de compréhension pragmatique (voir la *figure 3-10*), cela signifie que ces différentes informations pourraient être différemment pondérées selon les situations mais aussi selon les individus, ou différemment appelées activement. Dans notre cas, ces stratégies semblent liées à l'âge chronologique, mais d'autres caractéristiques ou traits pourraient être associés à de telles différences, telles que le neurotype (*chapitre 3, chapitre 6*) ou d'autres traits individuels ou situationnels. Certains travaux rapportent par exemple que la théorie de l'esprit pourrait être recrutée de façon flexible selon la situation (voir par exemple l'étude de Trott & Bergen [2019] sur les requêtes indirectes décrite au *chapitre 1*). Dans le domaine de la résolution de références en situation d'interaction, Apperly (2018) propose ainsi que « *la probabilité que les états mentaux d'autrui soient inférés, stockés ou utilisés dépend des dispositions des participants, de leurs motivations et de leurs ressources cognitives* » (p. 138). Nos résultats confortent donc cette hypothèse plus générale. Les différences liées à l'âge pourraient naturellement être liées aux ressources cognitives des participants, tant en termes de théorie de l'esprit que de langage, ainsi qu'à des aspects motivationnels ou tempéramentaux qu'on sait évoluer avec l'âge, tel que le rapport à l'incertitude ou à l'ambiguïté (Robinson & Whittaker, 1985). Ces considérations rejoignent également l'hypothèse

de Kissine (2015), qui proposait que différentes procédures permettent d'inférer l'intention de communication d'un locuteur, et que leur recrutement ou non dépendrait d'une compétence de *contrôle métacognitif*. Ainsi, des différences selon le neurotype entre participants autistes et non-autistes pourraient également être révélatrices de différences interindividuelles plus générales, et expliquées en termes de ressources cognitives, mais également de préférences cognitives ou de motivation, par exemple. Dans la partie suivante, nous examinerons plus en détail ces différences, et ce que nos données peuvent apporter aux modèles de l'autisme.

### 3. Implications pour les modèles de l'autisme

#### 3.1. Circonscrire les particularités pragmatiques de l'autisme

Nous avons développé au sein du *troisième chapitre* la tension qui apparaît entre des représentations de l'autisme intrinsèquement associées à des difficultés dans les inférences pragmatiques, et une littérature expérimentale qui nuance, et contredit même parfois ce constat. Les résultats de la *troisième étude* rapportée dans cette thèse, bien qu'inattendus à certains égards, s'inscrivent tout à fait dans ce mouvement, puisqu'ils mettent en évidence un contraste entre des phénomènes pragmatiques qui semblent appréhendés de façon typique par les participants autistes que nous avons évalués, et d'autres sur lesquels des difficultés apparaissent. Ces résultats renforcent donc la nécessité de préciser la nature des particularités pragmatiques associées à l'autisme, qui ne peuvent pas être décrites comme un déficit pragmatique global.

Deux facteurs importants émergent de la littérature comme pouvant circonscrire ces particularités. Le premier, non-spécifique à l'autisme, est la présence de fragilités verbales formelles qui sont fréquemment associées à cette condition (mais aussi en dehors de l'autisme, voir par exemple Durrleman & Delage [2016]), et qui pourraient être à l'origine de difficultés constatées de longue date. Ce facteur a notamment été mobilisé pour expliquer que certaines personnes autistes puissent être en difficulté et pas d'autre, par exemple avec les métaphores (e.g., Norbury, 2005), ou que des particularités autistiques apparaissent dans certaines tâches mais pas toutes (e.g., Kalandadze et al., 2019). Le deuxième facteur mobilisé est la demande en théorie de l'esprit qui serait associée à certaines situations pragmatiques mais pas à d'autres, et qui a conduit à la dichotomie entre pragmatique linguistique et pragmatique sociale (voir Andrés-Roqueta & Katsos, 2017; Kissine, 2021), en lien avec les particularités de théorie de l'esprit associées à l'autisme. Nous avons vu plus haut dans ce chapitre que nos résultats sont difficilement compatibles avec cette dichotomie en l'état, sans pour autant remettre en question l'impact de la théorie de l'esprit sur les performances pragmatiques, en général et dans l'autisme.

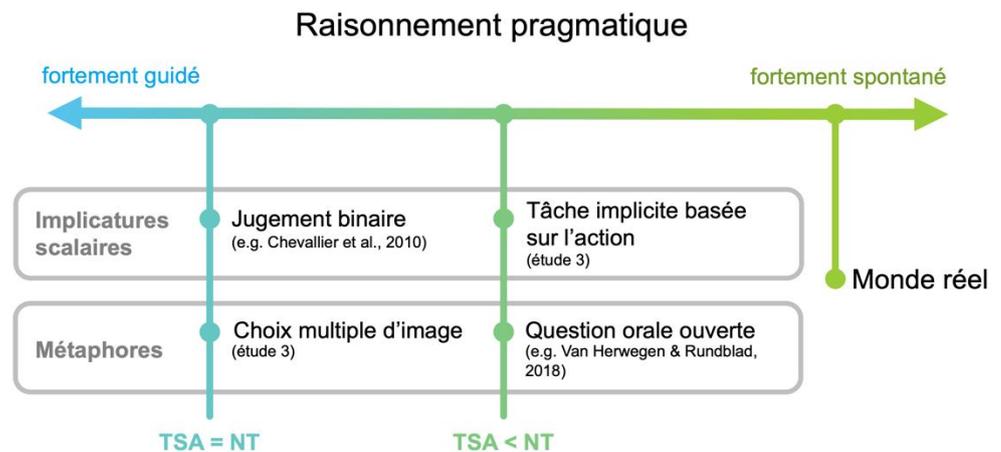
Ces deux facteurs n'apparaissent toutefois pas suffisants pour expliquer les résultats de notre troisième étude. Premièrement, le groupe d'enfants autistes que nous avons

recruté présentait des performances verbales qui correspondaient à ce qui était attendu pour son âge, et les deux tâches utilisées limitaient la demande langagière et métalinguistique en mobilisant du matériel verbal simple, fréquent, largement accessible et contrôlé, et des modes de réponse non-verbaux. Deuxièmement, si le statut des métaphores au sein de la dichotomie entre pragmatique linguistique et pragmatique sociale est incertain (particulièrement pour la tâche que nous avons utilisée, comme discuté *ci-dessus*), les implicatures scalaires sont typiquement considérées comme un cas de pragmatique linguistique. Or, c'est avec ce phénomène plutôt qu'avec les métaphores que les enfants autistes se sont montrés plus en difficulté, contrairement à ce qui était attendu. Il semble donc que nos résultats nécessitent la mobilisation d'un facteur supplémentaire pour expliquer le pattern de performance des participants autistes. La spontanéité des processus mobilisés, développée en *partie 2*, pourrait être une hypothèse pertinente. Il est intéressant de constater que cette possibilité était évoquée dès les deux premières études sur les implicatures scalaires dans l'autisme (Chevallier et al., 2010; Pijnacker et al., 2009). Chevallier et al. (2010) évoquaient ainsi : « *Une question importante pour la recherche à venir sera de déterminer dans quelle mesure ces résultats s'appliquent dans des contextes plus écologiques, d'autant plus que certains auteurs ont démontré que les individus autistes réalisent des performances différentes dans des dispositifs expérimentaux explicites et dans des situations plus écologiques* » (p. 1114, traduction personnelle).

À la différence de notre tâche d'implicatures scalaires, dont nous avons décrit plus haut le caractère implicite, et donc dépendant de processus spontanés, la tâche de métaphore du TIPi semble guider plus fortement le raisonnement pragmatique des participants, via des questions explicites et surtout des propositions de réponses sous forme d'images. Ces propositions font apparaître une interprétation métaphorique, qui n'a donc pas besoin d'être spontanément générée, et la compétition avec une interprétation littérale est ainsi 'organisée' par le dispositif. Les propositions de réponse font donc émerger des possibilités d'interprétation pertinentes, tout en réduisant le champ des possibles car seules trois interprétations (dont deux sont réellement en compétition) sont proposées. Ainsi, sans que cette tâche ne puisse être considérée comme trop facile puisqu'à l'inverse d'un effet plancher les données montrent une variabilité importante dans cette fenêtre d'âge, elle peut être décrite comme *guidant* le raisonnement pragmatique. On pourrait donc raisonnablement émettre l'hypothèse que les enfants autistes manifestent des difficultés à mettre *spontanément* en œuvre ce raisonnement, mais qu'ils soient en mesure de le faire lorsque la situation les guide pour déterminer l'interprétation la plus probable d'un énoncé. Cette hypothèse peut a posteriori s'appliquer à nos résultats mais aussi à d'autres résultats rapportés dans la littérature (figure 7-4). Dans le domaine des métaphores par exemple, les études qui mettent en évidence les différences les plus importantes entre participants autistes et non-autistes emploient notamment des questions ouvertes (e.g. Rundblad & Annaz, 2010; Van Herwegen & Rundblad, 2018). Ces tâches ont parfois été assimilées à des tâches d'explications verbales libres afin

d'expliquer les difficultés qu'elles causent manifestement aux participants autistes par leur charge verbale et métalinguistique (Kalandadze et al., 2019). Cependant cette catégorisation paraît très discutable : répondre à une question ouverte à l'aide d'un seul mot mobilise beaucoup moins les capacités verbales et socio-cognitives qu'expliquer le sens d'une expression à un interlocuteur. Regrouper ces deux modalités de réponse paraît donc réducteur. À cet égard, l'hypothèse de difficultés associées à la mise en place *spontanée* d'un raisonnement pragmatique pourrait efficacement expliquer qu'une question ouverte entraîne plus de difficultés qu'une question avec des propositions de réponses imagées comme dans notre tâche.

En outre, cette dichotomie permettrait de rendre compte de l'écart entre les effets observés dans des tâches expérimentales en laboratoire, pas toujours significatifs, sinon souvent faibles ou modérés, et d'autre part les difficultés rapportées par les personnes concernées, et les aidants familiaux et professionnels qui les accompagnent (*chapitre 3*). Ce contraste révèle que la plupart des tâches expérimentales échouent très probablement à capturer des processus à l'œuvre dans la vie quotidienne des personnes. L'identifier devrait constituer un objectif important des recherches ultérieures. L'hypothèse selon laquelle cet écart pourrait (en partie) être expliqué par une propension moindre à recruter et à déployer spontanément des inférences pragmatiques, plutôt que leur intégrité propre, semble ainsi pertinente à explorer.



**Figure 7-4 :** Positionnement des tâches de cette thèse et d'autres tâches de la littérature sur un axe opposant des raisonnements pragmatiques guidés vs spontanés, et impactant les performances de participants autistes (TSA) relativement à des contrôles neurotypiques (NT)

Cette hypothèse ne paraît pas incompatible avec les deux facteurs évoqués précédemment, la charge langagière et la demande en théorie de l'esprit, pouvant aussi influencer les performances des participants de façon orthogonale à ces deux caractéristiques, en particulier la charge langagière et métalinguistique de la situation. Alternativement, la spontanéité du recrutement des processus pragmatiques pourrait également être associée à la charge en théorie de l'esprit d'une tâche ou d'une situation. En effet, si ce raisonnement peut être guidé par des éléments explicites d'un paradigme expérimental – tels que des propositions de réponses – en leur absence ce raisonnement

pourrait être guidé par des attentes riches, basées sur une représentation des états mentaux d'autrui, précise et mise à jour continuellement. Pour ce qui est des métaphores, par exemple, dans un effort d'intégrer les difficultés des personnes autistes dans le cadre de la théorie de la pertinence, Wearing (2010) proposait ainsi que « *les informations sur les états mentaux du locuteur sont vitales pour GUIDER les directions particulières dans lesquelles les processus [...] peuvent opérer* » (p. 209, traduction et emphase personnelles).

Cette hypothèse s'appuie en outre sur des travaux réalisés dans d'autres domaines. De façon générale, il a par exemple été montré que des participants autistes avaient moins tendance à extraire *spontanément* les régularités de leur environnement, mais qu'ils pouvaient le faire autant que des participants contrôles neurotypiques lorsque cela leur était demandé explicitement (Sapey-Triomphe et al., 2018). Plus proche de notre sujet d'étude, comme évoqué ci-dessus et au [chapitre 3](#), dans le domaine de la théorie de l'esprit un écart entre des performances dans des tâches expérimentales et leur déploiement spontané dans la vie quotidienne a été décrit dans l'autisme (Klin et al., 2000, 2003; Senju, 2012). Cette hypothèse fait plus généralement écho à la théorie de la motivation sociale de l'autisme, qui propose « *que le TSA implique un défaut d'intérêt spontané pour mobiliser les compétences socio-cognitives dans des buts sociaux, mais que les compétences cognitives sous-jacentes pourraient être plus préservées qu'on ne le pensait précédemment* » (Chevallier et al., 2012, p. 237, traduction personnelle). De façon cohérente, une limite bien identifiée des interventions proposées dans l'autisme est une difficulté à généraliser dans la vie quotidienne des comportements ou compétences en dehors des contextes dans lesquels elles ont été stimulées (Rao et al., 2008).

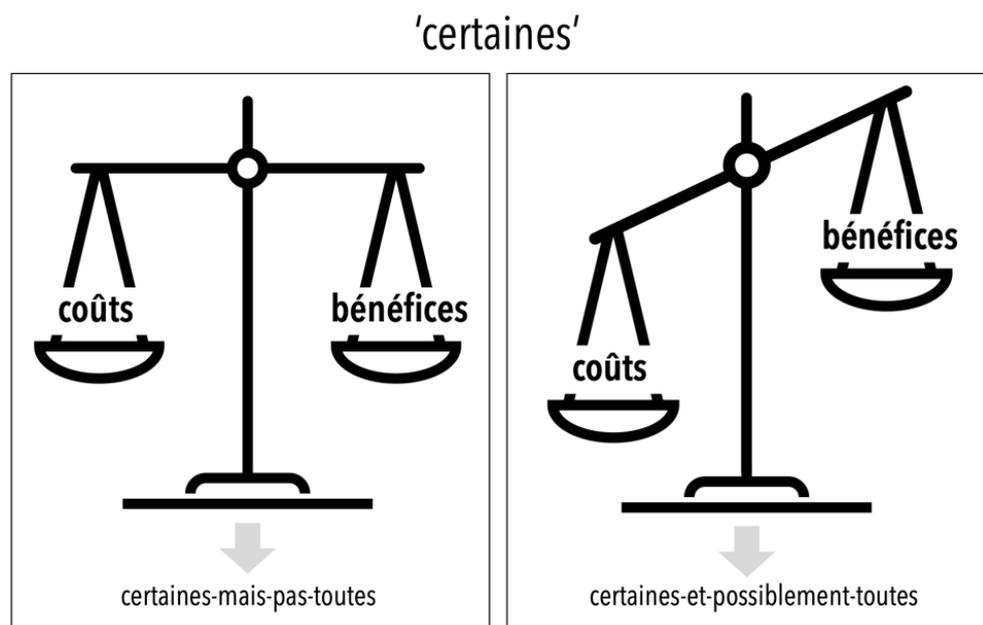
Plus spécifiquement dans le domaine de la compréhension pragmatique, cette hypothèse trouve également un soutien dans d'autres résultats expérimentaux. Dans une étude sur l'ironie, Zalla et al. (2014) ont par exemple montré que des adultes autistes pouvaient, comme un groupe contrôle neurotypique, inférer l'intention véhiculée par un énoncé ironique, et percevoir que certains métiers (par exemple, acteur) étaient associés à des probabilités plus importantes de produire un énoncé ironique que d'autres (par exemple, vétérinaire). En revanche, si les participants neurotypiques *utilisaient* ces stéréotypes pour moduler leur interprétation d'énoncés ironiques ou littéraux – interprétant par exemple un énoncé ironique comme étant plus moqueur dans un contexte où il était énoncé par un acteur que par un vétérinaire –, ce n'était pas le cas des participants autistes chez qui une telle modulation n'était pas observée. Ainsi, malgré des compétences et connaissances initiales comparables, *l'utilisation* de ces informations variées pour inférer l'intention de communication d'un locuteur différait selon les groupes.

Dans une autre étude, A.C. Wislon et Bishop (2021) ont par ailleurs évalué la compréhension d'implicatures variées chez des adultes autistes et non-autistes appariés en langage formel, montrant que les participants autistes sélectionnaient 5

fois plus fréquemment que les autres une proposition de réponse ‘*Je ne sais pas*’ lorsqu’elle était disponible. Toutefois, lorsque cette proposition n’était pas disponible, ils sélectionnaient le plus souvent la réponse correcte. Ce résultat suggère également que des propositions de réponse peuvent *guider* le raisonnement pragmatique de ces participants, sans quoi une tendance à demeurer sur une interprétation incertaine était constatée. Ce type de méthodologie constitue une perspective très appropriée pour évaluer l’hypothèse que nous proposons, en comparant les performances de participants autistes et non-autistes, sur des phénomènes pragmatiques particuliers, en manipulant expérimentalement le caractère guidant ou non de la situation, indépendamment de sa charge langagière et métalinguistique. Cette approche gagnerait également, dans des travaux ultérieurs, à adopter une approche transdiagnostique en incluant par exemple des participants classiquement catégorisés comme présentant un trouble du déficit de l’attention (TDAH), ou un trouble du langage. Ceci permettrait d’une part d’isoler les particularités autistiques d’autres particularités cliniques éventuelles, et d’autre part d’approcher la réalité clinique des personnes concernées, dans laquelle les frontières entre catégories diagnostiques sont floues et en partie artificielles (Astle et al., 2022), et où les comorbidités sont la norme plutôt qu’une exception (Levy et al., 2010). Actuellement, des expérimentations avec le TIPi sont ainsi en cours auprès d’enfants avec TDAH (trouble du déficit de l’attention / hyperactivité), cette condition étant également associée à des difficultés pragmatiques (Caillies et al., 2014; Carruthers et al., 2022), dont la nature et la distinction de celles associées à l’autisme doivent encore être précisées.

Une autre direction de recherche qui apparaît à l’issue de ce travail est le cadre que peut fournir la théorie de la pertinence pour décrire la nature ou l’origine des particularités pragmatiques associées à l’autisme. La pertinence est en effet définie comme une balance entre le coût du traitement d’une information et les effets positifs ou bénéfiques qu’elles apportent (*chapitre 1*). De nombreux travaux se sont focalisés, explicitement ou non, sur le coût de ces opérations dans l’autisme par rapport au développement typique, notamment en termes de ressources cognitives (mobilisation de la théorie de l’esprit ou des ressources langagières par exemple). Les résultats d’A.C. Wislon et Bishop (2021) évoqués ci-dessus poussent également à envisager ce coût sous un abord émotionnel et préférentiel, les auteurs invoquant une préférence pour la certitude (ou une intolérance à l’incertitude, voir Hwang et al. [2020]), et pas seulement une compétence pragmatique limitée pour expliquer les différences associées à l’autisme. Wilson et Bishop proposent que ces deux particularités s’alimenteraient mutuellement au cours du développement ; une difficulté à comprendre les significations implicites ou ambiguës aurait des conséquences sociales négatives, qui entraîneraient à leur tour une préférence pour la certitude, et réciproquement. Mais au-delà des coûts de ces opérations, il apparaît également que des difficultés à mettre en place des inférences pragmatiques (particulièrement s’il s’agit d’un défaut de recrutement spontané d’un raisonnement fonctionnel) pourraient résulter d’une évaluation atypique des *bénéfices* associés à cette information, déséquilibrant tout autant le rapport de pertinence (figure 7-5). En pratique clinique, il

n'est ainsi pas rare que les personnes autistes questionnent l'utilité même d'une communication implicite ou indirecte, comme en témoignent également les éléments rapportés au [chapitre 3](#). La théorie de la motivation sociale de l'autisme, en particulier, pousse à considérer que les informations inférées pragmatiquement pourraient être moins valorisées dans l'autisme que dans le développement typique. Développer des paradigmes permettant de manipuler ou de mesurer distinctement les coûts et les bénéfices attendus ou réels d'une inférence pragmatique permettrait ainsi probablement d'éclairer les atypies associées à l'autisme, en les distinguant (ou non) de celles qui concernent d'autres conditions cliniques rencontrant également des difficultés pragmatiques.



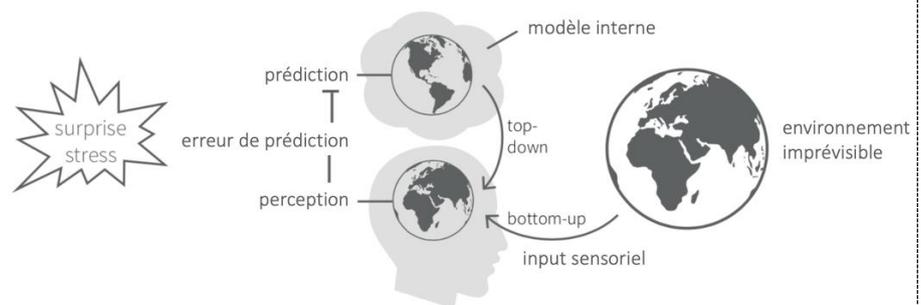
**Figure 7-5 :** Rapport entre les coûts et les bénéfices estimés d'une information, suite par exemple à la présentation du quantifieur 'quelques'. À gauche, le rapport est équilibré et pousse à inférer les intentions de communication déterminant l'énoncé (ici, en comprenant l'implicature scalaire). À droite, le rapport est déséquilibré et ne guide pas l'inférence, ce qui peut résulter d'une surévaluation des coûts et/ou d'une sous-estimations des bénéfices

En lien avec le problème de double empathie (Milton, 2012), il serait en outre pertinent de tester si des différences dans l'évaluation des bénéfices et des coûts d'un traitement pragmatique liées à l'autisme sont modulées, voire expliquées par des effets de groupes internes et externes. En d'autres termes, les coûts pourraient être surévalués, ou les bénéfices sous-évalués, dans le cas d'une information provenant d'un groupe externe plutôt qu'interne, ce qui pourrait expliquer au moins une partie des difficultés classiquement attribuées à l'autisme seul. Afin d'étudier ces questions, inclure des personnes autistes dans la conception et la mise en place des recherches, dans une démarche *participative*, devrait être privilégié afin de produire des résultats plus robustes, moins biaisés et plus directement interprétables (Marocchini, 2023).

Des perspectives intéressantes s’ouvrent également avec les modèles bayésiens<sup>22</sup> de la cognition en général, et de l’autisme en particulier, qui ont été développés au cours de la dernière décennie (voir Angeletos Chrysaitis & Seriès, 2023; Haker et al., 2016; Palmer et al., 2017). En effet les modèles que nous avons évoqués jusqu’à présent étaient principalement centrés sur ses particularités *sociales* de l’autisme, négligeant leur articulation avec les atypies *non-sociales* qui le caractérisent également, probablement de façon plus spécifique, notamment au niveau perceptif et comportemental. En outre, ces modèles proposent aussi une mise en lien concrète entre les niveaux comportementaux, cognitifs et biologiques souvent étudiés isolément dans l’autisme. Nous développerons cette perspective plus large dans l’*encart 4*, avant de nous tourner vers les implications et perspectives que cette thèse ouvre pour la pratique clinique dans la *partie suivante*.

**Encart 4** : Perspectives ouvertes par les modèles bayésiens de l’autisme

Les approches bayésiennes englobent différentes approches de la cognition qui considèrent le cerveau-esprit humain comme un système optimisé pour inférer un modèle interne du monde, à partir d’informations de l’environnement toujours parcellaires (Knill & Pouget, 2004). Elles ont particulièrement été appliquées à la perception (Petzschner et al., 2015) ou à l’intégration multisensorielle (Ernst & Banks, 2002). Selon ces approches, ces inférences sur le monde sont opérées en prédisant continuellement l’environnement à partir d’un modèle interne, puis en comparant ces prédictions aux informations issues des sens afin d’établir une perception de l’environnement (figure 7-6). La comparaison de ces perceptions aux prédictions aboutit ainsi à des erreurs de prédictions plus ou moins importantes qui peuvent être exploitées.



**Figure 7-6** : Représentation du principe de base des modèles bayésiens, extraite et traduite de Haker et al. (2016), représentant la comparaison des prédictions d’un modèle interne à l’input sensoriel issu de l’environnement et aboutissant à une perception et à des erreurs de prédiction que le système vise à minimiser.

Ce système gouvernerait la perception et l’action pour maintenir une homéostasie en minimisant les erreurs de prédictions. Schématiquement, deux moyens peuvent être employés pour ce faire : soit en *agissant* pour que la perception corresponde aux

<sup>22</sup> Nous incluons sous ce terme les modèles bayésiens au sens strict et les modèles de codage prédictif, qui se distinguent sur des aspects qui ne sont pas capitaux pour cette discussion.

prédictions (par exemple, en orientant différemment les récepteurs sensoriels), soit en *améliorant le modèle interne*, en le mettant à jour sur la base des informations perceptives (apprentissage). Cependant, toutes les erreurs de prédiction ne sont pas uniformément utilisées pour mettre à jour le modèle interne, puisqu'une large part d'entre-elles proviennent de bruits non-significatifs de l'environnement ou des récepteurs sensoriels. Les erreurs de prédiction sont ainsi modulées par l'incertitude quant à l'input sensoriel, et quant à la prédiction initiale, incertitudes elles-mêmes modélisées par ce système. Différentes représentations sont ainsi construites et enchâssées dans des structures hiérarchiques, des plus concrètes aux plus abstraites, avec des représentations de plus haut niveau basées sur des représentations de plus bas niveau. Ainsi, plus les représentations de haut niveau sont précises et spécifiées, moins elles seront altérées par des expériences surprenantes. Un intérêt particulier de ces approches computationnelles de la cognition est qu'elles permettent des prédictions précises et numériques, testables expérimentalement.

Elles se développent ainsi en psychiatrie (Stephan & Mathys, 2014), et notamment dans l'autisme, qui a typiquement été modélisé comme étant caractérisé par un déséquilibre entre des prédictions sous-spécifiées et des inputs sensoriels surreprésentés (ce qui peut être causé par différentes atypies), ce qui aboutit à des perceptions dominées par les informations sensorielles 'brutes', et qui entravent ainsi la formation de représentations de plus haut niveau (e.g., Pellicano & Burr, 2012). L'intérêt de ce modèle est qu'il permet de comprendre de façon cohérente différentes manifestations de l'autisme jusqu'alors peu conciliées ; dans le domaine perceptif par exemple, cela expliquerait un style perceptif plus focalisé sur les détails, discriminant plus difficilement les informations pertinentes des informations non-pertinentes du contexte, car moins guidées par des attentes de plus haut niveau, ce sur quoi se focalisait jusqu'alors la théorie de la faible cohérence centrale (Happé & Frith, 2006). Dans le domaine social, les interactions sont caractérisées par une forte complexité et une forte variabilité des situations qui sont en plus dynamiques, nécessitant ainsi des représentations de très haut niveau, plutôt qu'une accumulation de représentations de bas-niveau ; des difficultés à former des représentations de plus haut niveau aboutiraient ainsi à des difficultés de compréhension de ces situations, menant à des comportements socialement maladaptés, relativement à des personnes neurotypiques, et à des parasitages par des informations non-pertinentes (e.g. Robic et al., 2015), ce que rapportent fréquemment les personnes concernées. Un tel fonctionnement aboutirait également à des apprentissages ancrés dans les contextes dans lesquels ils ont eu lieu, mais peu généralisables à des situations différant même légèrement (nous reviendrons sur cette caractéristique de l'autisme dans la [section 2.2](#)). Enfin, l'ensemble de ces particularités serait à l'origine d'erreurs de prédictions amples et fréquentes, et donc d'un stress cognitif important que des mises à jour du modèle interne ne peuvent pas absorber. Une solution alternative pour rétablir l'homéostasie du système serait alors d'*agir* pour créer soi-même des stimulations et perceptions correspondant aux prédictions, ce qui se traduirait par des comportements répétitifs, stéréotypés, et des intérêts dits restreints, qui constituent un autre critère diagnostique de l'autisme.

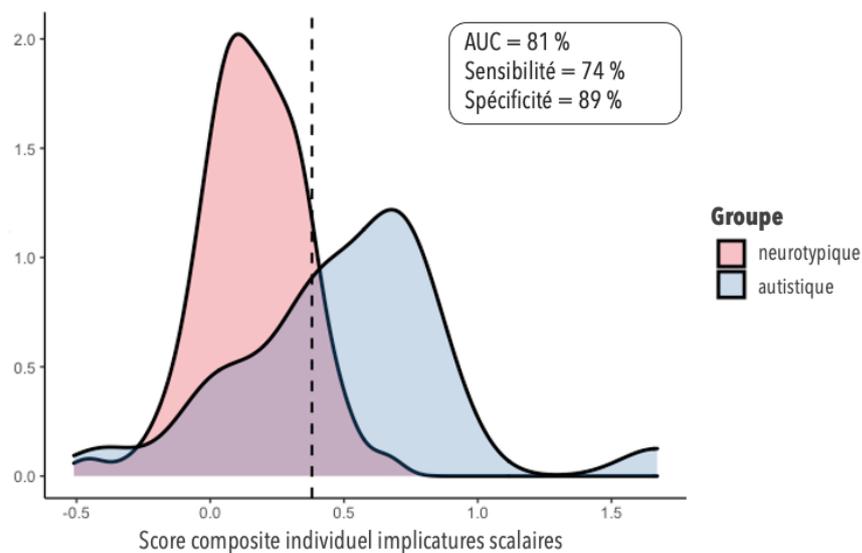
Même si les données disponibles sont encore peu consensuelles (voir la revue systématique d'Angeletos Chrysaitis & Seriès, 2023), ce cadre intégrant les manifestations sociales et non-sociales de l'autisme, ainsi que ses caractéristiques comportementales, cognitives et biologiques, mérite ainsi de plus amples investigations. S'il a été appliqué à des stimuli sociaux tels que le regard ou les visages, il ne l'a pas été à des stimuli pragmatiques tels que ceux manipulés dans cette thèse, à notre connaissance.

Or, certains modèles ont parallèlement été développés en pragmatique, tels que le RSA (voir Degen, 2023 et [chapitre 1](#)) qui propose une modélisation computationnelle de l'interprétation (littérale ou non) du langage, dans lequel l'intention de communication d'un locuteur est inférée selon un raisonnement bayésien à partir d'un input constitué par le décodage linguistique d'un énoncé. Ces outils et approches pourraient ainsi être combinés pour l'étude des particularités pragmatiques de l'autisme, permettant d'isoler plusieurs sources possibles de difficultés telles que les prédictions antérieures, le décodage littéral de l'information, ou l'intégration en elle-même de ces informations dans des modélisations récursives. De plus, ces approches permettraient d'étudier directement le lien entre les particularités pragmatiques et les particularités non-sociales (par exemple, perceptives) observées dans l'autisme. Enfin, les approches bayésiennes de l'autisme constituent des modèles dont l'application en clinique pourrait améliorer les pratiques de diagnostic et d'intervention (Haker et al., 2016).

### 3.2. Implications et perspectives pour la pratique clinique

Cette thèse nourrit également un certain nombre d'implications plus ou moins directes pour la pratique clinique. Nous examinerons dans un premier temps celles qui s'appliquent aux pratiques d'évaluation, avant de nous tourner vers l'intervention et l'accompagnement des personnes concernées.

Le projet de recherche qui supporte cette thèse vise en effet à valider pour la pratique clinique un nouvel outil d'évaluation, le TIPi (Test des Inférences Pragmatiques informatisé). La tâche d'implicatures scalaires montrant des différences inter-groupes nettes pourrait ainsi être particulièrement utile pour évaluer l'utilisation spontanée de l'informativité pour inférer les intentions de communication d'autrui, à des fins fonctionnelles mais également diagnostiques. Sur le modèle des analyses réalisées dans l'[étude annexe](#), dans laquelle nous avons validé pour la clinique une tâche de théorie de l'esprit classique en recherche, notre tâche d'implicatures scalaires pourrait être utilisée pour des évaluations individuelles. Des analyses réalisées à partir des données cliniques préliminaires du [chapitre 6](#) montrent ainsi qu'il est possible de calculer pour chaque participant un indice indexant le coût en termes de temps de réponse qu'induisent les conditions sous-informatives, relativement aux situations contrôles. Cet index peut lui-même être combiné au taux de réponses pragmatiques de chaque enfant, pour fournir une mesure composite de l'effort associé aux implicatures scalaires. Ce composite a montré une bonne capacité à discriminer les enfants autistes des enfants à développement typiques (figure 7-7), la courbe ROC (Receiver Operating Characteristics) révélant un bon caractère discriminant, avec une aire sous la courbe de 81 % (IC à 95% = [.67-.93]) ; le seuil optimal de ce score composite est associé à une sensibilité de 74% et une spécificité de 89%, soit de bonnes voire très bonnes performances.



**Figure 7-7 :** Distribution pour chacun des groupes de l'étude 3 du score composite individuel indexant l'effort cognitif associé aux implicatures scalaires, et seuil optimal (ligne pointillée) discriminant les deux groupes, avec les valeurs associées de sensibilité et de spécificité

La tâche de métaphore apparaît par contraste peu pertinente à des fins diagnostiques, puisqu'elle ne discrimine pas les enfants selon leur neurotype de façon générale. En revanche, elle montre de bonnes qualités psychométriques par ailleurs (distribution des données, validité structurelle et externe, cohérence interne), et pourrait être utilisée pour évaluer la compréhension de métaphores dans la situation particulière qu'elle induit (que nous avons qualifiée de *guidée*). Certains enfants autistes se sont ainsi montrés en difficulté dans cette tâche également, ce qui pourrait orienter l'intervention vers un travail de ces habiletés en elles-mêmes, en situation guidée et facilitante, avant de stimuler leur mise en place spontanée, en rééducation puis en vie quotidienne. De plus, d'autres enfants avec un trouble du neurodéveloppement pourraient être en difficulté avec ces processus et pourraient bénéficier de cette évaluation. Par ailleurs, quatre autres phénomènes pragmatiques sont distinctement évalués dans le TIPi mais n'ont pas été développés dans cette thèse : les implicatures ad-hoc, les inférences conditionnelles, l'ironie, et les implicatures conversationnelles sur la conjonction 'et'. Comme pour les deux tâches détaillées dans cette thèse, chaque mesure comprend des items tests pouvant être comparés à des items contrôles appariés mais n'impliquant qu'une compréhension littérale, et permet l'enregistrement des réponses comme des temps de réponse des enfants. Une perspective qui paraît particulièrement pertinente pour la clinique, à l'issue de cette thèse, est d'étudier individuellement les propriétés de chacune de ces tâches, puis d'examiner les modalités dans lesquelles l'ensemble des indices qu'elles fournissent pourraient être combinés pour constituer des scores composites particulièrement discriminant, ou isolant certains mécanismes cognitifs à l'œuvre. Comme nous avons vu que les performances et les temps de réponses des items tests et contrôles de la tâche d'implicatures scalaires pouvaient être combinés dans un score composite unique, la même démarche pourrait par exemple être appliquée à des combinaisons inter-tâches. Si la dichotomie entre situations guidées et

spontanées venait à être confirmée, il serait ainsi possible de construire à partir des mesures du TIPi un score composite indexant la capacité à mettre en œuvre spontanément des inférences pragmatiques, relativement à leur intégrité telle qu'observée dans des situations guidées.

Plus généralement, nos données cliniques confirment que réaliser une performance normale dans une tâche particulière ne garantit pas l'intégrité du raisonnement pragmatique dans d'autres situations, notamment dans d'autres tâches expérimentales, et a fortiori dans la vie quotidienne. Comme discuté plus haut, il apparaît probable que la plupart des tâches standardisées échouent à capturer une partie d'un fonctionnement atypique associé à l'autisme, et à l'origine d'un handicap fonctionnel important. Dans le raisonnement clinique, qu'il soit diagnostique ou plus fonctionnel, des difficultés pragmatiques ne doivent donc pas être écartées à la vue d'une performance normale dans une tâche. Nos résultats encouragent plutôt à considérer un ensemble d'indices, et à intégrer aux résultats d'épreuves standardisées les rapports des personnes concernées et de leurs proches, notamment des parents et des enseignants pour ce qui est des enfants, que des questionnaires peuvent permettre d'objectiver et de quantifier. Le questionnaire développé par A.C. Wilson (2022) paraît à ce titre très pertinent, puisqu'à la différence d'autres outils tels que la CCC (Children's Communication Checklist, Bishop, 1998), construite à partir d'un cadre théorique, ce nouveau questionnaire a été développé et validé à partir des expériences rapportées par les personnes concernées elles-mêmes, ce qui explique sans doute qu'il se montre particulièrement discriminant. Une traduction française et une adaptation à un public pédiatrique constituent des perspectives pertinentes pour la clinique en France. Cela fournirait en outre un moyen de mettre en rapport quantitativement les observations réalisées en vie quotidienne et celles issues d'évaluations directes standardisées. Pour l'heure, l'observation clinique, outillée conceptuellement par les repères de fonctionnement et de développement typiques développés dans cette thèse, doit également garder une place importante dans l'évaluation clinique de la pragmatique, dans l'attente que des outils standardisés valides et fiables soient disponibles, ce vers quoi ce travail de thèse tend humblement.

Nos résultats illustrent également que les inférences pragmatiques s'appuient sur des bases cognitives multiples, au premier rang desquelles se trouvent les habiletés langagières formelles et la théorie de l'esprit, et que ces bases peuvent différer selon les situations, mais également selon les personnes (*étude 2*). En d'autres termes, il faut considérer qu'un individu peut probablement mobiliser différentes compétences cognitives pour réaliser une inférence pragmatique, et que des déficits différents pourraient entraver ces processus selon les individus. Ceci confirme l'importance en pratique clinique d'inscrire l'évaluation des habiletés pragmatiques dans une évaluation plus globale du fonctionnement, notamment langagier et cognitif, pour inférer les forces et les faiblesses d'un individu particulier. Ceci doit éclairer le fonctionnement particulier d'un individu et ainsi alimenter le raisonnement diagnostique, mais aussi orienter l'accompagnement dont il pourra bénéficier.

Nous avons en effet déjà évoqué dans les pages précédentes (*chapitre 3*) que les inférences pragmatiques d'enfants autistes pouvaient être développées, spontanément ou par certaines interventions cliniques ; cependant, la multiplicité des hypothèses sur la nature des particularités autistiques dans le traitement du langage, et plus encore sur l'origine de ces particularités, ne donne pas de directions claires pour orienter ces interventions, et très peu de données sont disponibles quant aux méthodes qui sont efficaces (Chahboun et al., 2021; Vulchanova et al., 2015; Vulchanova & Vulchanov, 2022). Pour ce qui est de la compréhension des métaphores, Melogno et Pinto (2022) évoquent notamment différentes approches (qu'ils qualifient de comportementales ou de psycholinguistiques selon leurs ancrages théoriques) qui consistent souvent à entraîner l'enfant à réaliser des analyses sémantiques des termes impliqués dans des métaphores, afin de guider le lien qui peut être établi entre les deux domaines différents. Autrement dit, ces approches mobilisent les habiletés sémantiques des participants pour guider les inférences pragmatiques. Parallèlement, Tonini et al. (2022) ont proposé avec le programme MetaCom un entraînement focalisé sur les aspects inférentiels et contextuels des métaphores, ancré dans la théorie de la pertinence. Appliqué en groupe à des enfants neurotypiques, cet entraînement a également montré un effet supérieur à une intervention contrôle, et une généralisation à la compréhension de texte. Ces différentes approches offrent donc autant de perspectives pour l'accompagnement des enfants autistes, mais très peu de preuves sont encore disponibles quant à l'approche à privilégier, chez quels enfants, quand, et à quelle intensité. Dans cette attente, nos résultats rappellent cette hétérogénéité et incitent à adapter l'approche (ré)éducative proposée à chaque enfant sur la base d'une évaluation globale, en ciblant ses faiblesses et en exploitant ses forces relatives. À l'avenir, comparer des interventions basées sur le langage ou sur les processus inférentiels, recrutant éventuellement la théorie de l'esprit, chez des enfants avec différents profils cliniques, constitue ainsi une perspective riche de recherche clinique, qui renseignerait également sur la direction et la causalité des liens entre inférences pragmatiques, habiletés langagières formelles et théorie de l'esprit.

L'hypothèse de difficultés spécifiques à mettre en œuvre *spontanément* des inférences pragmatiques, par rapport à des situations plus guidantes, rappelle également les difficultés à généraliser les apprentissages fréquemment observées dans l'autisme (Rao et al., 2008), comme évoqué ci-dessus. Si cette hypothèse demande des confirmations expérimentales, elle pousse également à faire de la généralisation en vie quotidienne un objectif essentiel des accompagnements proposés. Pour ce faire, les interventions dans le domaine de la communication sociale peuvent impliquer les parents, montrant alors des effets supérieurs à celles ciblant les enfants seulement (Parsons et al., 2017), ou impliquer des pairs neurotypiques (Watkins et al., 2015). Elles incluent également fréquemment des exercices à faire dans le quotidien, en dehors des séances d'intervention (Parsons et al., 2017). Ces orientations doivent permettre de favoriser l'utilisation spontanée des inférences pragmatiques dans les interventions proposées. Il semble également pertinent de baser les interventions sur des situations concrètes vécues par les personnes concernées, afin d'une part de

s'approcher des difficultés précises que rencontre un individu (que les évaluations ne peuvent qu'estimer approximativement), et d'autre part des contextes spécifiques dans lesquelles elles se manifestent, favorisant l'application en vie quotidienne. Des outils tels que des journaux de bord des difficultés de communication rencontrées, alimentées par les jeunes concernées et/ou leurs proches pourraient faciliter ce travail. Afin de s'approcher de conditions écologiques, citons enfin les technologies de réalité virtuelle, qui constituent une perspective intéressante pour l'intervention, dans la mesure où elles permettent de reconstituer des contextes variés et relativement naturels, tout en offrant un environnement contrôlé, reproductible et modulable, et ainsi sécurisant, avant la mise en pratique en situation réelle (Aresti-Bartolome & Garcia-Zapirain, 2014; Berenguer et al., 2020). En France, les orthophonistes semblent d'ailleurs prêts à se saisir de ces technologies (Moron-Grand-Dufay, 2023).

Finalement, deux négligences de la littérature scientifique comme de cette thèse constituent d'autres perspectives importantes pour le développement des connaissances d'une part, et pour l'amélioration des pratiques cliniques d'autre part. Premièrement, les compétences pragmatiques sont très largement étudiées sur le versant de la compréhension, mais très peu de la production (voir par exemple Melogno & Pinto, 2022), probablement car cela constitue un défi méthodologique plus important encore que l'étude de la compréhension. Or, '*savoir quoi dire*' constitue un élément central des difficultés de communication vécues par les personnes autistes (A. C. Wilson, 2022), ce qui devrait en faire une priorité de recherche. Deuxièmement, comme cela a déjà été évoqué précédemment, une large partie de la recherche est focalisée sur l'autisme (et plus encore sur des formes d'autisme relativement isolées, sans comorbidités, à la différence de ses présentations dans le monde réel), mais l'autisme ne constitue ni la seule condition associée à des difficultés pragmatiques, ni même la plus prévalente, eu égard aux troubles du langage, de l'attention, des apprentissages, ou à d'autres conditions psychiatriques, pour ce qui est du seul champ pédiatrique. Si la conduite de recherches dans l'autisme permet de faire avancer la connaissance et les pratiques de façon générale, et bénéficie ainsi secondairement à d'autres populations, les spécificités de différentes catégories diagnostiques, ou les variations interindividuelles associées à différents traits cliniques, doivent également faire l'objet de recherches spécifiques.

## Conclusion

Une large part de notre communication verbale repose sur des significations implicites. Les inférences pragmatiques, qui permettent d'inférer ce qu'un locuteur *voulait dire* en disant ce qu'il a dit, sont donc une compétence clef pour naviguer dans notre environnement social. Pour mesurer leur intégrité dans un contexte clinique pédiatrique, nous proposons un Test des Inférences Pragmatiques informatisé (TIPi), dont deux nouvelles tâches évaluant les implicatures scalaires et les métaphores ont fait l'objet de cette thèse.

Alors que des travaux récents ont en mis évidence des précurseurs précoces de ces inférences à l'âge préscolaire, nos résultats soutiennent que la fenêtre de 6 à 11 ans reste une importante période de développement de ces habiletés. Ce développement est notamment visible pour les *implicatures scalaires*, avec lesquelles nous mettons en évidence une propension grandissante à utiliser spontanément l'informativité pour inférer les intentions de communication d'un locuteur. De même, les performances des enfants en *compréhension de références métaphoriques* progressent également fortement à cet âge. Nos résultats révèlent en outre une bascule dans les prédicteurs cognitifs associés aux performances, de la théorie de l'esprit chez les enfants les plus jeunes aux compétences langagières formelles chez les plus âgés, ce qui suggère un changement qualitatif et pas seulement quantitatif dans le raisonnement pragmatique. De plus, nos observations confirment que ce développement à l'âge scolaire est également actif chez les enfants autistes, mais avec des trajectoires qui sont toutefois atypiques par rapport aux enfants neurotypiques. Plutôt que de révéler une incapacité pragmatique globale, ou un retard généralisé, nos données dévoilent un pattern de particularités spécifiques, avec des atypies autistiques visibles ici sur les implicatures scalaires, mais pas sur les métaphores, contrairement à ce qui était attendu.

Ces contributions expérimentales mettent en évidence les apports mutuels que peuvent s'offrir la pragmatique expérimentale et la clinique du neurodéveloppement. D'une part, l'observation des différences interindividuelles, et notamment de fonctionnements atypiques, met à l'épreuve les modèles pragmatiques existants. C'est ainsi que le concept de *spontanéité* du raisonnement pragmatique nous a semblé mériter d'être considéré, pour éclairer nos résultats et d'autres faits établis dans la littérature scientifique. Nos données questionnent également la dichotomie entre pragmatique linguistique et sociale, qui se montre intéressante mais insuffisante face aux observations que nous avons rapportées. Réciproquement, le déploiement de concepts et de méthodes précises issues de la pragmatique expérimentale permet des observations détaillées du fonctionnement autistique, qui aident à circonscrire les caractéristiques de cette condition. Ces connaissances constituent des outils conceptuels qui permettent aux cliniciens de mieux appréhender les expériences des

populations qu'ils reçoivent, et donc de mieux les accompagner ; elles doivent aussi aider la société en général à mieux les comprendre. Plus concrètement, ces apports ont également permis le développement d'outils d'évaluation pour les professionnels de santé ; en fournissant une estimation plus précise des compétences des enfants accueillis, ils doivent favoriser le repérage et l'orientation vers des accompagnements adaptés ceux d'entre eux qui paraissent particulièrement vulnérables. Ce travail illustre ainsi qu'entre l'essentialisation de nos différences et leur négation, appréhender ce qui nous distingue éclaire finalement ce qui nous rassemble, nous permettant de mieux nous comprendre, mieux nous accepter, mieux nous entraider, dans l'espoir d'une société plus inclusive et respectueuse de chacun.

Nous sommes des ténors du silence. Habités dans nos tripes par l'onde de choc des cerveaux sans contrôle.

Avoir des envies d'exister dans son identité, c'est un combat et j'accepte.

J'accepte de me battre contre ce moi déconnant aux yeux du monde.

J'accepte de me battre pour ce moi déconnant aux yeux du monde.

Mais, je refuse de me battre pour les yeux du monde.

BABUILLEC, AUTISTE  
SANS PAROLES (2016)

# RÉFÉRENCES

## Liste des références bibliographiques

*N.B. Cette liste inclut les références introduites dans les différents chapitres de cette thèse, à l'exception des articles (chapitres 4 et 5, annexes B et D), dont les références sont présentées directement à l'issue du texte.*

Abell, F., Happe, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15*(1), 1–16.

Ackerman, B. P. (1981). When is a question not answered? The understanding of young children of utterances violating or conforming to the rules of conversational sequencing. *Journal of Experimental Child Psychology, 31*, 487–507. [https://doi.org/10.1016/0022-0965\(81\)90032-1](https://doi.org/10.1016/0022-0965(81)90032-1)

Airenti, G. (2017). Pragmatic Development. In L. Cummings (Ed.), *Research in Clinical Pragmatics* (pp. 3–28). Springer International Publishing. [https://doi.org/10.1007/978-3-319-47489-2\\_1](https://doi.org/10.1007/978-3-319-47489-2_1)

Amaral, D. G., Schumann, C. M., & Nordahl, C. W. (2008). Neuroanatomy of autism. *Trends in Neurosciences, 31*(3), 137–145. <https://doi.org/10.1016/j.tins.2007.12.005>

American Psychiatric Association. (2022). *Diagnostic And Statistical Manual Of Mental Disorders: DSM-5-TR* (Fifth Edition, Text Revision (DSM-5-TR)). American Psychiatric Association.

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*, 996. <https://doi.org/10.3389/fpsyg.2017.00996>

Andrés-Roqueta, C., & Katsos, N. (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](https://doi.org/10.1044/2020_JSLHR-19-00263)

Angeletos Chrysaitis, N., & Seriès, P. (2023). 10 years of Bayesian theories of autism: A comprehensive review. *Neuroscience & Biobehavioral Reviews, 145*, 105022. <https://doi.org/10.1016/j.neubiorev.2022.105022>

Apperly, I. (2018). Mindreading and Psycholinguistic Approaches to Perspective Taking: Establishing Common Ground. *Topics in Cognitive Science*, 10(1), 133–139. <https://doi.org/10.1111/tops.12308>

Aresti-Bartolome, N., & Garcia-Zapirain, B. (2014). Technologies as support tools for persons with autistic spectrum disorder: A systematic review. *International Journal of Environmental Research and Public Health*, 11(8), 7767–7802.

Arzouan, Y., Goldstein, A., & Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research*, 1160, 69–81.

Asch, S., & Nerlove, H. (1960). The development of double function terms in children: An exploratory investigation. In B. Kaplan & S. Wapner, *Perspectives in psychological theory: Essays in honor of Heinz Werner* (pp. 47–60). International Universities Press.

Asperger, H. (1944). Die „Autistischen psychopathen“ im Kindesalter (U. Frith, Trans.). *Archiv Für Psychiatrie Und Nervenkrankheiten*, 117(1), 37–92.

Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2022). Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry*, 63(4), 397–417. <https://doi.org/10.1111/jcpp.13481>

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, 62(5), 875–890.

Baltazar, M., Geoffray, M., Chatham, C., Bouvard, M., Martinez Teruel, A., Monnet, D., Scheid, I., Murzi, E., Couffin-Cadiergues, S., Umbricht, D., Murtagh, L., Delorme, R., Ly Le-Moal, M., Leboyer, M., & Amestoy, A. (2021). “Reading the Mind in the Eyes” in Autistic Adults is Modulated by Valence and Difficulty: An InFoR Study. *Autism Research*, 14(2), 380–388. <https://doi.org/10.1002/aur.2390>

Bambini, V., Agostoni, G., Buonocore, M., Tonini, E., Bechi, M., Ferri, I., Sapienza, J., Martini, F., Cuoco, F., Cocchi, F., Bischetti, L., Cavallaro, R., & Bosia, M. (2022). It is time to address language disorders in schizophrenia: A RCT on the efficacy of a novel training targeting the pragmatics of communication (PragmaCom). *Journal of Communication Disorders*, 97, 106196. <https://doi.org/10.1016/j.jcomdis.2022.106196>

Bambini, V., Arcara, G., Bechi, M., Buonocore, M., Cavallaro, R., & Bosia, M. (2016). The communicative impairment as a core feature of schizophrenia: Frequency of pragmatic deficit, cognitive substrates, and relation with quality of life. *Comprehensive Psychiatry*, 71, 106–120. <https://doi.org/10.1016/j.comppsy.2016.08.012>

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016).

Disentangling Metaphor from Context: An ERP Study. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00559>

Bannard, C., & Tomasello, M. (2012). Can we dissociate contingency learning from social learning in word acquisition by 24-month-olds? *PloS One*, 7(11), e49881.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, 118(1), 84–93. <https://doi.org/10.1016/j.cognition.2010.10.010>

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

Baron-Cohen, S. (2010). Empathizing, systemizing, and the extreme male brain theory of autism. In *Progress in Brain Research* (Vol. 186, pp. 167–175). Elsevier. <https://doi.org/10.1016/B978-0-444-53630-3.00011-7>

Baron-Cohen, S., 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](https://doi.org/10.1016/0010-0277(85)90022-8)

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *J. Child Psychol. Psychiat.*, 42(2), 241–251.

Baxter, A. J., Brugha, T. S., Erskine, H. E., Scheurer, R. W., Vos, T., & Scott, J. G. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine*, 45(3), 601–613. <https://doi.org/10.1017/S003329171400172X>

Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic Review and Inventory of Theory of Mind Measures for Young Children. *Frontiers in Psychology*, 10, 2905. <https://doi.org/10.3389/fpsyg.2019.02905>

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology*, 41(2), 328.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, 8(6), 492–499.

Berenguer, C., Baixauli, I., Gómez, S., Andrés, M. de E. P., & De Stasio, S. (2020). Exploring the impact of augmented reality in children and adolescents with autism spectrum disorder: A systematic review. *International Journal of Environmental Research and Public Health*, 17(17), 6143.

- Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1450–1460. <https://doi.org/10.1037/a0027850>
- Birch, S. A., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, 107(3), 1018–1034.
- Bishop, D. V. M. (1998). Development of the Children's Communication Checklist (CCC): A Method for Assessing Qualitative Aspects of Communicative Impairment in Children. *Journal of Child Psychology and Psychiatry*, 39(6), 879–891. <https://doi.org/10.1017/S0021963098002832>
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Adams, C., Archibald, L., Baird, G., Bauer, A., Bellair, J., Boyle, C., Brownlie, E., Carter, G., Clark, B., Clegg, J., Cohen, N., Conti-Ramsden, G., Dockrell, J., Dunn, J., Ebbels, S., ... house, A. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 58(10), 1068–1080. <https://doi.org/10.1111/jcpp.12721>
- Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, 9(4), Article 4. <https://doi.org/10.1038/nrn2353>
- Bonnefon, J.-F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, 112(2), 249–258.
- Bosco, F. M., Gabbatore, I., & Tirassa, M. (2014). A broad assessment of theory of mind in adolescence: The complexity of mindreading. *Consciousness and Cognition*, 24, 84–97. <https://doi.org/10.1016/j.concog.2014.01.003>
- 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>
- Botha, M., Hanlon, J., & Williams, G. L. (2021). Does Language Matter? Identity-First Versus Person-First Language Use in Autism Research: A Response to Vivanti. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-020-04858-w>
- Bott, L., & Noveck, I. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51(3), 437–457. <https://doi.org/10.1016/j.jml.2004.05.006>
- Bowdle, B. F., & Gentner, D. (2005). The Career of Metaphor. *Psychological Review*, 112(1), 193–216. <https://doi.org/10.1037/0033-295X.112.1.193>

Bowes, A., & Katz, A. (2015). Metaphor creates intimacy and temporarily enhances theory of mind. *Memory & Cognition*, 43(6), 953–963. <https://doi.org/10.3758/s13421-015-0508-4>

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, 100(3), 434–463. <https://doi.org/10.1016/j.cognition.2005.07.003>

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge university press.

Buijsman, R., Begeer, S., & Scheeren, A. M. (2023). ‘Autistic person’ or ‘person with autism’? Person-first language preference in Dutch adults with autism and parents. *Autism*, 27(3), 788–795. <https://doi.org/10.1177/13623613221117914>

Bury, S. M., Jellett, R., Spoor, J. R., & Hedley, D. (2020). “It defines who I am” or “It’s something I have”: What language do [autistic] Australian adults [on the autism spectrum] prefer? *Journal of Autism and Developmental Disorders*, 1–11.

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective Imitation of In-Group Over Out-Group Members in 14-Month-Old Infants. *Child Development*, 84(2), 422–428. <https://doi.org/10.1111/j.1467-8624.2012.01860.x>

Cable.co.uk. (2022). *Woldwide Speed League Data* [dataset].

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

Canal, P., & Bambini, V. (2023). Pragmatics Electrified. In M. Grimaldi, Y. Shtyrov, & E. Brattico, *Language Electrified. Techniques, Methods, Applications, and Future Perspectives in the Neurophysiological Investigation of Language*. Springer. <https://doi.org/10.31234/osf.io/5z7b2>

Canal, P., Bischetti, L., Bertini, C., Ricci, I., Lecce, S., & Bambini, V. (2022). N400 differences between physical and mental metaphors: The role of Theories of Mind. *Brain and Cognition*, 161, 105879. <https://doi.org/10.1016/j.bandc.2022.105879>

Cappelli, G., Noccetti, S., Arcara, G., & Bambini, V. (2018). Pragmatic competence and its relationship with the linguistic and cognitive profile of young adults with dyslexia. *Dyslexia*, 24(3), 294–306. <https://doi.org/10.1002/dys.1588>

Carlson, P., & Anisfeld, M. (1969). Some Observations on the Linguistic Competence of a Two-Year-Old Child. *Child Development*, 40(2), 569–575.

<https://doi.org/10.2307/1127424>

Carruthers, S., Taylor, L., Sadiq, H., & Tripp, G. (2022). The profile of pragmatic language impairments in children with ADHD: A systematic review. *Development and Psychopathology*, 34(5), 1938–1960. <https://doi.org/10.1017/S0954579421000328>

Carston, R. (2008). *Thoughts and utterances: The pragmatics of explicit communication*. John Wiley & Sons.

Carston, R., & Wearing, C. (2011). Metaphor, hyperbole and simile: A pragmatic approach. *Language and Cognition*, 3(2), 283–312. <https://doi.org/10.1515/langcog.2011.010>

Chahboun, S., Kvello, Ø., & Page, A. G. (2021). Extending the Field of Extended Language: A Literature Review on Figurative Language Processing in Neurodevelopmental Disorders. *Frontiers in Communication*, 6, 661528. <https://doi.org/10.3389/fcomm.2021.661528>

Chahboun, S., Vulchanov, V., Saldaña, D., Eshuis, H., & Vulchanova, M. (2016). Can You Play with Fire and Not Hurt Yourself? A Comparative Study in Figurative Language Comprehension between Individuals with and without Autism Spectrum Disorder. *PloS One*, 11(12), e0168571. <https://doi.org/10.1371/journal.pone.0168571>

Champagne-Lavau, M., & Joannette, Y. (2009). Pragmatics, theory of mind and executive functions after a right-hemisphere lesion: Different patterns of deficits. *Journal of Neurolinguistics*, 22(5), 413–426.

Champagne-Lavau, M., & Stip, E. (2010). Pragmatic and executive dysfunction in schizophrenia. *Journal of Neurolinguistics*, 23(3), 285–296. <https://doi.org/10.1016/j.jneuroling.2009.08.009>

Chapman, R. (2019). Neurodiversity theory and its discontents: Autism, schizophrenia, and the social model of disability. *The Bloomsbury Companion to Philosophy of Psychiatry*, 371.

Chevallier, C., Kohls, G., Troiani, V., Brodtkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, 16(4), 231–239. <https://doi.org/10.1016/j.tics.2012.02.007>

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar Inferences in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 40(9), 1104–1117. <https://doi.org/10.1007/s10803-010-0960-8>

Chierchia, G., Guasti, M. T., Gualmini, A., Meroni, L., Crain, S., & Foppolo, F. (2004). Semantic and pragmatic competence in children's and adults' comprehension of or. *Experimental Pragmatics*, 283–300.

- Chouinard, B., & Cummine, J. (2016). All the world's a stage: Evaluation of two stages of metaphor comprehension in people with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 23, 107–121. <https://doi.org/10.1016/j.rasd.2015.12.008>
- Chouinard, B., Volden, J., Hollinger, J., & Cummine, J. (2019). Spoken metaphor comprehension: Evaluation using the metaphor interference effect. *Discourse Processes*, 56(3), 270–287. <https://doi.org/10.1080/0163853X.2018.1455166>
- Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, 17(2), 417–431.
- Clark, H. H. (1996). *Using language*. Cambridge University Press.
- Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, 5(9), eaaw2594. <https://doi.org/10.1126/sciadv.aaw2594>
- Crompton, C. J., Ropar, D., Evans-Williams, C. V., Flynn, E. G., & Fletcher-Watson, S. (2020). *Autistic peer-to-peer information transfer is highly effective*.
- Csibra, G. (2010). Recognizing Communicative Intentions in Infancy. *Mind & Language*, 25(2), 141–168. <https://doi.org/10.1111/j.1468-0017.2009.01384.x>
- Cummings, L. (2016). *Research in clinical pragmatics*. Springer Berlin Heidelberg.
- Cummings, L. (Ed.). (2021). *Handbook of Pragmatic Language Disorders: Complex and Underserved Populations*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-74985-9>
- De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, 54(2), 128–133.
- de Villiers, J. (2007). The interface of language and Theory of Mind. *Lingua*, 117(11), 1858–1878. <https://doi.org/10.1016/j.lingua.2006.11.006>
- Deckert, M., Schmoeger, M., Schaunig-Busch, I., & Willinger, U. (2019). Metaphor processing in middle childhood and at the transition to early adolescence: The role of chronological age, mental age, and verbal intelligence. *Journal of Child Language*, 46(2), 334–367. <https://doi.org/10.1017/S0305000918000491>
- Degen, J. (2023). The Rational Speech Act Framework. *Annual Review of Linguistics*, 9(1), 519–540. <https://doi.org/10.1146/annurev-linguistics-031220-010811>
- Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, 39(4), 667–710.

- Del Sette, P., Bambini, V., Bischetti, L., & Lecce, S. (2020). Longitudinal associations between theory of mind and metaphor understanding during middle childhood. *Cognitive Development, 56*, 100958. <https://doi.org/10.1016/j.cogdev.2020.100958>
- Del Sette, P., Ronchi, L., Bambini, V., & Lecce, S. (2021). Longitudinal associations between metaphor understanding and peer relationships in middle childhood. *Infant and Child Development, 30*(4). <https://doi.org/10.1002/icd.2232>
- Deliens, G., Antoniou, K., Clin, E., & Kissine, M. (2017). Perspective-taking and frugal strategies: Evidence from sarcasm detection. *Journal of Pragmatics, 119*, 33–45. <https://doi.org/10.1016/j.pragma.2017.08.002>
- 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>
- Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences, 1*(4), 568–570. <https://doi.org/10.1017/S0140525X00076664>
- Dews, S., Kaplan, J., & Winner, E. (1995). Why not say it directly? The social functions of irony. *Discourse Processes, 19*(3), 347–367.
- Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using fiction to assess mental state understanding: A new task for assessing theory of mind in adults. *PloS One, 8*(11), e81279.
- Durrleman, S., Burnel, M., De Villiers, J. G., Thommen, E., Yan, R., & Delage, H. (2019). The Impact of Grammar on Mentalizing: A Training Study Including Children With Autism Spectrum Disorder and Developmental Language Disorder. *Frontiers in Psychology, 10*, 2478. <https://doi.org/10.3389/fpsyg.2019.02478>
- Durrleman, S., & Delage, H. (2016). Autism Spectrum Disorder and Specific Language Impairment: Overlaps in Syntactic Profiles. *Language Acquisition, 23*(4), 361–386. <https://doi.org/10.1080/10489223.2016.1179741>
- Durrleman, S., Peristeri, E., & Tsimpli, I. M. (2022). The language-communication divide: Evidence from bilingual children with atypical development. *Evolutionary Linguistic Theory, 4*(1), 5–51. <https://doi.org/10.1075/elt.00037.dur>
- Edey, R., Cook, J., Brewer, R., Johnson, M. H., Bird, G., & Press, C. (2016). Interaction takes two: Typical adults exhibit mind-blindness towards those with autism spectrum disorder. *Journal of Abnormal Psychology, 125*, 879–885. <https://doi.org/10.1037/abn0000199>
- Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., Montiel-

Nava, C., Patel, V., Paula, C. S., Wang, C., Yasamy, M. T., & Fombonne, E. (2012). Global Prevalence of Autism and Other Pervasive Developmental Disorders: Global epidemiology of autism. *Autism Research*, 5(3), 160–179. <https://doi.org/10.1002/aur.239>

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433.

Fairchild, S., & Papafragou, A. (2021). The Role of Executive Function and Theory of Mind in Pragmatic Computations. *Cognitive Science*, 45(2). <https://doi.org/10.1111/cogs.12938>

Falkum, I. L., & Köder, F. (2020). The acquisition of figurative meanings. *Journal of Pragmatics*, 164, 18–24. <https://doi.org/10.1016/j.pragma.2020.04.007>

Filippova, E. (2014). Irony production and comprehension. In *Pragmatic development in first language acquisition* (pp. 261–278). John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.10.15fil>

Filippova, E., & Astington, J. W. (2008). Further Development in Social Reasoning Revealed in Discourse Irony Understanding. *Child Development*, 79(1), 126–138. <https://doi.org/10.1111/j.1467-8624.2007.01115.x>

Fletcher-Watson, S., & Happé, F. (2019). *Autism: A new introduction to psychological theory and current debate*. Routledge.

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar Implicatures in Child Language: Give Children a Chance. *Language Learning and Development*, 8(4), 365–394. <https://doi.org/10.1080/15475441.2011.626386>

Francés, L., Quintero, J., Fernández, A., Ruiz, A., Caules, J., Fillon, G., Hervás, A., & Soler, C. V. (2022). Current state of knowledge on the prevalence of neurodevelopmental disorders in childhood according to the DSM-5: A systematic review in accordance with the PRISMA criteria. *Child and Adolescent Psychiatry and Mental Health*, 16(1), 27. <https://doi.org/10.1186/s13034-022-00462-1>

Frith, U. (1989). A new look at language and communication in autism. *International Journal of Language & Communication Disorders*, 24(2), 123–150.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431), 459–473. <https://doi.org/10.1098/rstb.2002.1218>

Frith, U., & Happé, F. (1994). Autism: Beyond “theory of mind.” *Cognition*, 50(1–3), 115–132.

Gao, Z., Yang, Q., Ma, X., Becker, B., Li, K., Zhou, F., & Kendrick, K. M. (2017).

Men Who Compliment a Woman's Appearance Using Metaphorical Language: Associations with Creativity, Masculinity, Intelligence and Attractiveness. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.02185>

Geelhand, P., Papastamou, F., Belenger, M., Clin, E., Hickman, L., Keating, C. T., & Sowden, S. (2023). Autism-related language preferences of French-speaking autistic adults: An online survey. *Autism in Adulthood*.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, 47–59.

Gentner, D., & Clement, C. (1988). Evidence for relational selectivity in the interpretation of analogy and metaphor. *Psychology of Learning and Motivation*, 22, 307–358.

Georgieff, N. (2017). Les paradoxes de l'autisme: Limites des modèles et théories, perspectives de recherche. *Perspectives Psy*, 4, 308–319.

Gernsbacher, M. A., & Pripas-Kapit, S. R. (2012). Who's Missing the Point? A Commentary on Claims that Autistic Persons Have a Specific Deficit in Figurative Language Comprehension. *Metaphor and Symbol*, 27(1), 93–105. <https://doi.org/10.1080/10926488.2012.656255>

Geschwind, D. H. (2008). Autism: Many Genes, Common Pathways? *Cell*, 135(3), 391–395. <https://doi.org/10.1016/j.cell.2008.10.016>

Geurts, B., Kissine, M., & Tiel, B. van. (2019). Pragmatic reasoning in autism. In K. Morsanyi & R. M. J. Byrne (Eds.), *Thinking, Reasoning, and Decision Making in Autism* (1st ed., pp. 113–134). Routledge. <https://doi.org/10.4324/9781351060912-6>

Gibbs, R. W. (1990). Comprehending figurative referential descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 56–66. <https://doi.org/10.1037/0278-7393.16.1.56>

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

Gibbs, R. W. (2000). Irony in Talk Among Friends. *Metaphor and Symbol*, 15(1–2), 5–27. <https://doi.org/10.1080/10926488.2000.9678862>

Gibbs, R. W. (2008). *The Cambridge handbook of metaphor and thought*. Cambridge University Press.

Glucksberg, S., Gildea, P., & Bookin, H. B. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, 21(1), 85–98.

- Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, *97*(1), 3.
- Gold, R., & Faust, M. (2010). Right hemisphere dysfunction and metaphor comprehension in young adults with Asperger syndrome. *Journal of Autism and Developmental Disorders*, *40*(7), 800–811. <https://doi.org/10.1007/s10803-009-0930-1>
- Gold, R., Faust, M., & Goldstein, A. (2010). Semantic integration during metaphor comprehension in Asperger syndrome. *Brain and Language*, *113*(3), 124–134. <https://doi.org/10.1016/j.bandl.2010.03.002>
- Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. <https://doi.org/10.1016/j.tics.2016.08.005>
- Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and Implicature: Modeling Language Understanding as Social Cognition. *Topics in Cognitive Science*, *5*(1), 173–184. <https://doi.org/10.1111/tops.12007>
- Gottman, J., Gonso, J., & Rasmussen, B. (1975). Social Interaction, Social Competence, and Friendship in Children. *Child Development*, *76*, 709–718.
- Grassmann, S. (2014). The pragmatics of word learning. In *Pragmatic development in first language acquisition* (pp. 139–160). John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.10.09gra>
- Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, *13*(1), 252–263.
- Grice, P. (1969). Utterer's Meaning and Intention. *The Philosophical Review*, *78*(2), 147. <https://doi.org/10.2307/2184179>
- Grice, P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.
- Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.
- Grosse, G., Behne, T., Carpenter, M., & Tomasello, M. (2010). Infants communicate in order to be understood. *Developmental Psychology*, *46*(6), 1710–1722. <https://doi.org/10.1037/a0020727>
- Guasti, T. M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667–696. <https://doi.org/10.1080/01690960444000250>
- Habibe, J. (2023, April 21). 6 powerful lessons from running a marathon during your

PhD. *The Struggling Scientists 2023*. <https://thestrugglingscientists.com/marathon-lessons/>

Haker, H., Schneebeli, M., & Stephan, K. E. (2016). Can Bayesian Theories of Autism Spectrum Disorder Help Improve Clinical Practice? *Frontiers in Psychiatry*, 7. <https://doi.org/10.3389/fpsy.2016.00107>

Hancock, J. T. (2004). Verbal Irony Use in Face-To-Face and Computer-Mediated Conversations. *Journal of Language and Social Psychology*, 23(4), 447–463. <https://doi.org/10.1177/0261927X04269587>

Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48(2), 101–119.

Happé, F. (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. *Journal of Autism and Developmental Disorders*, 24(2), 129–154. <https://doi.org/10.1007/BF02172093>

Happé, F., Cook, J. L., & Bird, G. (2017). The Structure of Social Cognition: In(ter)dependence of Sociocognitive Processes. *Annual Review of Psychology*, 68(1), 243–267. <https://doi.org/10.1146/annurev-psych-010416-044046>

Happé, F., & Frith, U. (2006). The Weak Coherence Account: Detail-focused Cognitive Style in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 36(1), 5–25. <https://doi.org/10.1007/s10803-005-0039-0>

Harris, P. (1982). Cognitive prerequisites to language? *British Journal of Psychology*, 73(2), 187–195.

Hartshorne, J. K., Snedeker, J., Liem Azar, S. Y.-M., & Kim, A. E. (2015). The neural computation of scalar implicature. *Language, Cognition and Neuroscience*, 30(5), 620–634. <https://doi.org/10.1080/23273798.2014.981195>

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*, S001094522300045X. <https://doi.org/10.1016/j.cortex.2023.01.013>

Helland, W. A., Lundervold, A. J., Heimann, M., & Posserud, M.-B. (2014). Stable associations between behavioral problems and language impairments across childhood – The importance of pragmatic language problems. *Research in Developmental Disabilities*, 35(5), 943–951. <https://doi.org/10.1016/j.ridd.2014.02.016>

Hermann, I., Haser, V., van Elst, L. T., Ebert, D., Müller-Feldmeth, D., Riedel, A., & Konieczny, L. (2013). Automatic metaphor processing in adults with Asperger

syndrome: A metaphor interference effect task. *European Archives of Psychiatry and Clinical Neuroscience*, 263 Suppl 2, S177-187. <https://doi.org/10.1007/s00406-013-0453-9>

Hirata, I., Mohri, I., Kato-Nishimura, K., Tachibana, M., Kuwada, A., Kagitani-Shimono, K., Ohno, Y., Ozono, K., & Taniike, M. (2016). Sleep problems are more frequent and associated with problematic behaviors in preschoolers with autism spectrum disorder. *Research in Developmental Disabilities*, 49, 86–99.

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>

Hoicka, E. (2014). The pragmatic development of humor. In *Pragmatic development in first language acquisition* (pp. 219–238). John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.10.13hoi>

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC press.

Holyoak, K. J., & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin*, 144(6), 641–671. <https://doi.org/10.1037/bul0000145>

Horn, L. R. (1989). *A natural history of negation* (University of Chicago Press).

Horton, W. S. (2007). Metaphor and readers' attributions of intimacy. *Memory & Cognition*, 35(1), 87–94. <https://doi.org/10.3758/BF03195945>

Houston-Price, C., Goddard, K., Séclier, C., Grant, S. C., Reid, C. J., Boyden, L. E., & Williams, R. (2011). Tracking speakers' false beliefs: Is theory of mind available earlier for word learning? *Developmental Science*, 14(4), 623–634.

Huang, Y. (2017). *The Oxford Handbook of Pragmatics*. Oxford University Press.

Huang, Y. T., & Snedeker, J. (2009a). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, 58(3), 376–415. <https://doi.org/10.1016/j.cogpsych.2008.09.001>

Huang, Y. T., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, 45(6), 1723–1739. <https://doi.org/10.1037/a0016704>

Hwang, Y. I., Arnold, S., Srasuebkul, P., & Trollor, J. (2020). Understanding anxiety in adults on the autism spectrum: An investigation of its relationship with intolerance of uncertainty, sensory sensitivities and repetitive behaviours. *Autism*, 24(2), 411–422.

Jary, M. (2013). Two types of implicature: Material and behavioural. *Mind & Language*, 28(5), 638–660.

Kalandadze, T., Bambini, V., & Næss, K.-A. B. (2019). A systematic review and meta-analysis of studies on metaphor comprehension in individuals with autism spectrum disorder: Do task properties matter? *Applied Psycholinguistics*, 40(6), 1421–1454. <https://doi.org/10.1017/S0142716419000328>

Kalandadze, T., Norbury, C., Naerland, T., & Naess, K.-A. B. (2018). Figurative Language Comprehension in Individuals with Autism Spectrum Disorder: A Meta-Analytic Review. *Autism: The International Journal of Research and Practice*, 22(2), 99–117. <https://doi.org/10.1177/1362361316668652>

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2(3), 217–250.

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, 111(33), 12002–12007.

Karmiloff-Smith, A. (1995). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, 2(10), 389–398. [https://doi.org/10.1016/s1364-6613\(98\)01230-3](https://doi.org/10.1016/s1364-6613(98)01230-3)

Katsos, N. (2014). Scalar implicature. *Pragmatic Development in First Language Acquisition*, 10, 183–197.

Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, 120(1), 67–81. <https://doi.org/10.1016/j.cognition.2011.02.015>

Katsos, N., Cummins, C., Ezeizabarrena, M.-J., Gavarró, A., Kuvač Kraljević, J., Hrzica, G., Grohmann, K. K., Skordí, A., Jensen de López, K., Sundahl, L., van Hout, A., Hollebrandse, B., Overweg, J., Faber, M., van Koert, M., Smith, N., Vija, M., Zupping, S., Kunnari, S., ... Noveck, I. (2016). Cross-linguistic patterns in the acquisition of quantifiers. *Proceedings of the National Academy of Sciences*, 113(33), 9244–9249. <https://doi.org/10.1073/pnas.1601341113>

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism*, 20(4), 442–462. <https://doi.org/10.1177/1362361315588200>

Kerbrat-Orecchioni, C. (1998). La notion d'interaction en linguistique: Origines,

apports, bilan. *Langue Française*, 51–67.

Khomsî, A., Khomsî, F., & Pasquet, F. (2007). *BILO - Bilans Informatisés de Langage Oral*. Pearson.

Kissine, M. (2012). Pragmatics, Cognitive Flexibility and Autism Spectrum Disorders. *Mind & Language*, 27(1), 1–28. <https://doi.org/10.1111/j.1468-0017.2011.01433.x>

Kissine, M. (2015). Pragmatics as Metacognitive Control. *Frontiers in Psychology*, 6, 2057. <https://doi.org/10.3389/fpsyg.2015.02057>

Kissine, M. (2021). Autism, constructionism, and nativism. *Language*, 97(3), e139–e160.

Kissine, M., & De Brabanter, P. (2023). Pragmatic responses to under-informative some-statements are not scalar implicatures. *Cognition*, 237, 105463. <https://doi.org/10.1016/j.cognition.2023.105463>

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1430), 345–360. <https://doi.org/10.1098/rstb.2002.1202>

Klin, A., Schultz, R., & Cohen, D. (2000). Theory of mind in action: Developmental perspectives on social neuroscience. *Understanding Other Minds: Perspectives from Developmental Neuroscience*, 2, 357–388.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12), 712–719.

Kogan, N., & Chadrow, M. (1986). Children's comprehension of metaphor in the pictorial and verbal modality. *International Journal of Behavioral Development*, 9(3), 285–295.

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, 1284, 145–155.

Lakoff, G., & Johnson, M. (1980). The Metaphorical Structure of the Human Conceptual System. *Cognitive Science*, 4(2), 195–208. [https://doi.org/10.1207/s15516709cog0402\\_4](https://doi.org/10.1207/s15516709cog0402_4)

Langdon, R., & Coltheart, M. (1999). Mentalising, schizotypy, and schizophrenia. *Cognition*, 71(1), 43–71. [https://doi.org/10.1016/S0010-0277\(99\)00018-9](https://doi.org/10.1016/S0010-0277(99)00018-9)

Law, J., Rush, R., Clegg, J., Peters, T., & Roulstone, S. (2015). The Role of Pragmatics in Mediating the Relationship Between Social Disadvantage and Adolescent Behavior. *Journal of Developmental & Behavioral Pediatrics*, 36(5), 389–398.

<https://doi.org/10.1097/DBP.0000000000000180>

Lecce, S., Ronchi, L., Del Sette, P., Bischetti, L., & Bambini, V. (2019). Interpreting physical and mental metaphors: Is Theory of Mind associated with pragmatics in middle childhood? *Journal of Child Language*, *46*(2), 393–407. <https://doi.org/10.1017/S030500091800048X>

Lee, G. T., Xu, S., Zou, H., Gilic, L., & Lee, M. W. (2019). Teaching children with autism to understand metaphors. *The Psychological Record*, *69*, 499–512.

Lee, J. J., & Pinker, S. (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review*, *117*(3), 785–807. <https://doi.org/10.1037/a0019688>

Leung, R. C., & Zakzanis, K. K. (2014). Brief Report: Cognitive Flexibility in Autism Spectrum Disorders: A Quantitative Review. *Journal of Autism and Developmental Disorders*, *44*(10), 2628–2645. <https://doi.org/10.1007/s10803-014-2136-4>

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Levy, S. E., Giarelli, E., Lee, L.-C., Schieve, L. A., Kirby, R. S., Cunniff, C., Nicholas, J., Reaven, J., & Rice, C. E. (2010). Autism Spectrum Disorder and Co-occurring Developmental, Psychiatric, and Medical Conditions Among Children in Multiple Populations of the United States. *Journal of Developmental & Behavioral Pediatrics*, *31*(4), 267. <https://doi.org/10.1097/DBP.0b013e3181d5d03b>

Littlemore, J. (2016). Metaphor use in educational contexts: Functions and variations. In *The Routledge handbook of metaphor and language* (pp. 301–313). Routledge.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Marocchini, E. (2023). Impairment or difference? The case of Theory of Mind abilities and pragmatic competence in the Autism Spectrum. *Applied Psycholinguistics*, 1–19. <https://doi.org/10.1017/S0142716423000024>

Marocchini, E., Di Paola, S., Mazzaggio, G., & Domaneschi, F. (2022). Understanding indirect requests for information in high-functioning autism. *Cognitive Processing*, *23*(1), 129–153. <https://doi.org/10.1007/s10339-021-01056-z>

Marr. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman and company.

Mashal, N., & Kasirer, A. (2011). Thinking Maps Enhance Metaphoric Competence in Children with Autism and Learning Disabilities. *Research in Developmental Disabilities: A Multidisciplinary Journal*, *32*(6), 2045–2054.

<https://doi.org/10.1016/j.ridd.2011.08.012>

Mathieson, F., Jordan, J., Carter, J. D., & Stubbe, M. (2016). Nailing Down Metaphors in CBT: Definition, Identification and Frequency. *Behavioural and Cognitive Psychotherapy*, *44*(2), 236–248. <https://doi.org/10.1017/S1352465815000156>

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual Differences in Children's Pragmatic Ability: A Review of Associations with Formal Language, Social Cognition, and Executive Functions. *Language Learning and Development*, *14*(3), 186–223. <https://doi.org/10.1080/15475441.2018.1455584>

Mazzaggio, G., Foppolo, F., Job, R., & Surian, L. (2021). Ad-hoc and scalar implicatures in children with autism spectrum disorder. *Journal of Communication Disorders*, *90*, 106089. <https://doi.org/10.1016/j.jcomdis.2021.106089>

Mazzarella, D. (2015). Politeness, relevance and scalar inferences. *Journal of Pragmatics*, *79*, 93–106. <https://doi.org/10.1016/j.pragma.2015.01.016>

Mazzarella, D., & Noveck, I. (2021). Pragmatics and mind reading: The puzzle of autism (Response to Kissine). *Language*, *97*(3), e198–e210. <https://doi.org/10.1353/lan.2021.0037>

Mazzarella, D., Reinecke, R., Noveck, I., & Mercier, H. (2018). Saying, presupposing and implicating: How pragmatics modulates commitment. *Journal of Pragmatics*, *133*, 15–27. <https://doi.org/10.1016/j.pragma.2018.05.009>

Mazzarella, D., Trouche, E., Mercier, H., & Noveck, I. (2018). Believing what you're told: Politeness and scalar inferences. *Frontiers in Psychology*, 908.

McElhanon, B. O., McCracken, C., Karpen, S., & Sharp, W. G. (2014). Gastrointestinal symptoms in autism spectrum disorder: A meta-analysis. *Pediatrics*, *133*(5), 872–883.

Melogno, S., & Pinto, M. A. (2022). Devising Trainings to Enhance the Capabilities of Children With Autism Spectrum Disorder to Cope With Metaphor: A Review of the Literature. *Frontiers in Communication*, *7*, 915873. <https://doi.org/10.3389/fcomm.2022.915873>

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838.

Miller, K., Schmitt, C., Chang, H.-H., & Munn, A. (2005). Young children understand some implicatures. *Proceedings of the 29th Annual Boston University Conference on Language Development*, 389–400.

Milton, D. E. M. (2012). On the ontological status of autism: The 'double empathy problem.' *Disability & Society*, *27*(6), 883–887.

<https://doi.org/10.1080/09687599.2012.710008>

Miranda, A., Berenguer, C., Roselló, B., & Baixauli, I. (2020). Relationships between the social communication questionnaire and pragmatic language, socialization skills, and behavioral problems in children with autism spectrum disorders. *Applied Neuropsychology: Child*, 9(2), 141–152. <https://doi.org/10.1080/21622965.2018.1550403>

Moon, E. (2002). *Speed of Dark*. Orbit.

Moron-Grand-Dufay, L. (2023). *Remédiation des habiletés sociales des adolescents de 12 à 20 ans présentant un Trouble du Spectre Autistique et technologies immersives (telles que la Réalité Virtuelle). Questionnaire de recensement des besoins en termes de situations sociales et de l'adhésion des professionnels pour l'emploi des technologies immersives dans leur pratique* [Mémoire d'orthophonie]. Université Claude Bernard Lyon 1.

Morsanyi, K., Hamilton, J., Stamenković, D., & Holyoak, K. J. (2022). Linking metaphor comprehension with analogical reasoning: Evidence from typical development and autism spectrum disorder. *British Journal of Psychology*, 113(2), 479–495. <https://doi.org/10.1111/bjop.12542>

Morsanyi, K., Stamenković, D., & Holyoak, K. J. (2020). Metaphor processing in autism: A systematic review and meta-analysis. *Developmental Review*, 57, 100925. <https://doi.org/10.1016/j.dr.2020.100925>

Müller, E., Schuler, A., & Yates, G. B. (2008). Social challenges and supports from the perspective of individuals with Asperger syndrome and other autism spectrum disabilities. *Autism*, 12(2), 173–190. <https://doi.org/10.1177/1362361307086664>

Murphy, S. M., Faulkner, D. M., & Farley, L. R. (2014). The Behaviour of Young Children with Social Communication Disorders During Dyadic Interaction with Peers. *Journal of Abnormal Child Psychology*, 42(2), 277–289. <https://doi.org/10.1007/s10802-013-9772-6>

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, 63(3), 324–346. <https://doi.org/10.1016/j.jml.2010.06.005>

Nilsen, E. S., Glenwright, M., & Huyder, V. (2011). Children and Adults Understand That Verbal Irony Interpretation Depends on Listener Knowledge. *Journal of Cognition and Development*, 12(3), 374–409. <https://doi.org/10.1080/15248372.2010.544693>

Nippold, M. A. (1998). *Later Language Development: The School-Age and Adolescent*

*Years. Second Edition. Pro-Ed, 8700 Shoal Creek Blvd.*

Norbury, C. F. (2004). Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech, Language, and Hearing Research: JSLHR*, 47(5), 1179–1193.

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>

Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language*, 41(3), 485–510.

Nordahl-Hansen, A., Tøndevold, M., & Fletcher-Watson, S. (2018). Mental health on screen: A DSM-5 dissection of portrayals of autism spectrum disorders in film and TV. *Psychiatry Research*, 262, 351–353. <https://doi.org/10.1016/j.psychres.2017.08.050>

Noveck, I. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78(2), 165–188. [https://doi.org/10.1016/S0010-0277\(00\)00114-1](https://doi.org/10.1016/S0010-0277(00)00114-1)

Noveck, I. (2018). *Experimental Pragmatics: The Making of a Cognitive Science*. Cambridge University Press. <https://doi.org/10.1017/9781316027073>

Noveck, I., Bianco, M., & Castry, A. (2001). The Costs and Benefits of Metaphor. *Metaphor and Symbol*, 16(1–2), 109–121. <https://doi.org/10.1080/10926488.2001.9678889>

O’Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A Clinical Assessment Tool for Advanced Theory of Mind Performance in 5 to 12 Year Olds. *Journal of Autism and Developmental Disorders*, 39(6), 916–928. <https://doi.org/10.1007/s10803-009-0699-2>

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308(5719), 255–258. <https://doi.org/10.1126/science.1107621>

O’Reilly, C., Lewis, J. D., & Elsabbagh, M. (2017). Is functional brain connectivity atypical in autism? A systematic review of EEG and MEG studies. *PLOS ONE*, 12(5), e0175870. <https://doi.org/10.1371/journal.pone.0175870>

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, 17(4), 465–477.

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism:

Towards volatility, action, and behavior. *Psychological Bulletin*, 143(5), 521–542. <https://doi.org/10.1037/bul0000097>

Papafragou, A., Friedberg, C., & Cohen, M. L. (2018). The Role of Speaker Knowledge in Children’s Pragmatic Inferences. *Child Development*, 89(5), 1642–1656. <https://doi.org/10.1111/cdev.12841>

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition*, 86(3), 253–282. [https://doi.org/10.1016/S0010-0277\(02\)00179-8](https://doi.org/10.1016/S0010-0277(02)00179-8)

Parsons, L., Cordier, R., Munro, N., Joosten, A., & Speyer, R. (2017). A systematic review of pragmatic language interventions for children with autism spectrum disorder. *PLOS ONE*, 12(4), e0172242. <https://doi.org/10.1371/journal.pone.0172242>

Pastor-Cerezuela, G., Tordera Yllescas, J. C., González-Sala, F., Montagut-Asunción, M., & Fernández-Andrés, M.-I. (2018). Comprehension of Generalized Conversational Implicatures by Children With and Without Autism Spectrum Disorder. *Frontiers in Psychology*, 9, 272. <https://doi.org/10.3389/fpsyg.2018.00272>

Pellicano, E., & Burr, D. (2012). When the world becomes ‘too real’: A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, 16(10), 504–510. <https://doi.org/10.1016/j.tics.2012.08.009>

Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that...” attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, 39(3), 437–471. [https://doi.org/10.1016/0022-0965\(85\)90051-7](https://doi.org/10.1016/0022-0965(85)90051-7)

Petranovich, C. L., Walz, N. C., Staat, M. A., Chiu, C.-Y. P., & Wade, S. L. (2017). Structural language, pragmatic communication, behavior, and social competence in children adopted internationally: A pilot study. *Applied Neuropsychology: Child*, 6(4), 315–326. <https://doi.org/10.1080/21622965.2016.1182433>

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, 19(5), 285–293.

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J.-P., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, 39(4), 607–618. <https://doi.org/10.1007/s10803-008-0661-8>

Pinker, S. (2007). The evolutionary social psychology of off-record indirect speech acts. *Intercultural Pragmatics*, 4(4). <https://doi.org/10.1515/IP.2007.023>

Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics*, 25, 51–79. <https://doi.org/10.1075/bjl.25.04pou>

Pouscoulous, N. (2014). The elevator's buttocks. *Metaphorical Abilities in Children*. In D. Matthews (Ed.), *Pragmatic Development in First Language Acquisition*, 239–259.

Pouscoulous, N., Noveck, I., Politzer, G., & Bastide, A. (2007). A Developmental Investigation of Processing Costs in Implicature Production. *Language Acquisition*, 14(4), 347–375. <https://doi.org/10.1080/10489220701600457>

Pouscoulous, N., & Tomasello, M. (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, 156, 160–167. <https://doi.org/10.1016/j.pragma.2019.05.021>

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515–526. <https://doi.org/10.1017/S0140525X00076512>

Prevost, M., Carrier, M.-E., Chowne, G., Zelkowitz, P., Joseph, L., & Gold, I. (2014). The Reading the Mind in the Eyes test: Validation of a French version and exploration of cultural variations in a multi-ethnic city. *Cognitive Neuropsychiatry*, 19(3), 189–204. <https://doi.org/10.1080/13546805.2013.823859>

Prior, M., & Hoffmann, W. (1990). Brief report: Neuropsychological testing of autistic children through an exploration with frontal lobe tests. *Journal of Autism and Developmental Disorders*.

Rao, P. A., Beidel, D. C., & Murray, M. J. (2008). Social Skills Interventions for Children with Asperger's Syndrome or High-Functioning Autism: A Review and Recommendations. *Journal of Autism and Developmental Disorders*, 38(2), 353–361. <https://doi.org/10.1007/s10803-007-0402-4>

Reboul, A. C. (2015). Why language really is not a communication system: A cognitive view of language evolution. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01434>

Recanati, F. (2004). *Literal meaning*. Cambridge University Press.

Reinhart, T. (2004). The processing cost of reference set computation: Acquisition of stress shift and focus. *Language Acquisition*, 12(2), 109–155.

Roberts, C. (1996). Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Semantics and Pragmatics*, 5.

Robic, S., Sonié, S., Fonlupt, P., Henaff, M.-A., Touil, N., Coricelli, G., Mattout, J., & Schmitz, C. (2015). Decision-Making in a Changing World: A Study in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 45(6), 1603–1613. <https://doi.org/10.1007/s10803-014-2311-7>

Robinson, E., & Whittaker, S. (1985). Children's responses to ambiguous messages and their understanding of ambiguity. *Developmental Psychology*, *21*(3), 446.

Ronderos, C. R., & Noveck, I. (2023). Slowdowns in scalar implicature processing: Isolating the intention-reading costs in the Bott & Noveck task. *Cognition*, *238*, 105480.

Rubio Fernandez, P. (2007). Suppression in Metaphor Interpretation: Differences between Meaning Selection and Meaning Construction. *Journal of Semantics*, *24*(4), 345–371. <https://doi.org/10.1093/jos/ffm006>

Rundblad, G., & Annaz, D. (2010). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism: The International Journal of Research and Practice*, *14*(1), 29–46. <https://doi.org/10.1177/1362361309340667>

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*(4), 1054–1070.

Sapey-Triomphe, L.-A., Sonié, S., Hénaff, M.-A., Mattout, J., & Schmitz, C. (2018). Adults with Autism Tend to Undermine the Hidden Environmental Structure: Evidence from a Visual Associative Learning Task. *Journal of Autism and Developmental Disorders*, *48*(9), 3061–3074. <https://doi.org/10.1007/s10803-018-3574-1>

Saylor, M. M., Sabbagh, M. A., & Baldwin, D. A. (2002). Children use whole-part juxtaposition as a pragmatic cue to word meaning. *Developmental Psychology*, *38*(6), 993.

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, *19*(2), 65–72. <https://doi.org/10.1016/j.tics.2014.11.007>

Schaeken, W., Van Haeren, M., & Bambini, V. (2018). The Understanding of Scalar Implicatures in Children With Autism Spectrum Disorder: Dichotomized Responses to Violations of Informativeness. *Frontiers in Psychology*, *9*, 1266. <https://doi.org/10.3389/fpsyg.2018.01266>

Schieve, L. A., Rice, C., Yeargin-Allsopp, M., Boyle, C. A., Kogan, M. D., Drews, C., & Devine, O. (2012). Parent-reported prevalence of autism spectrum disorders in US-born children: An assessment of changes within birth cohorts from the 2003 to the 2007 National Survey of Children's Health. *Maternal and Child Health Journal*, *16*, 151–157.

Schovanec, J., & Glorion, C. (2012). *Je suis à l'Est ! Savant et autiste, un témoignage unique* (Poche). Plon.

- Scott-Phillips, T. (2014). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Bloomsbury Publishing.
- Scott-Phillips, T. (2017). Pragmatics and the aims of language evolution. *Psychonomic Bulletin & Review*, 24(1), 186–189. <https://doi.org/10.3758/s13423-016-1061-2>
- Seigneuric, A., Megherbi, H., Bueno, S., Lebahar, J., & Bianco, M. (2016). Children's comprehension skill and the understanding of nominal metaphors. *Journal of Experimental Child Psychology*, 150, 346–363. <https://doi.org/10.1016/j.jecp.2016.06.008>
- Semino, E. (2014). Pragmatic failure, mind style and characterisation in fiction about autism. *Language and Literature (Harlow, England)*, 23(2), 141–158. <https://doi.org/10.1177/0963947014526312>
- Senju, A. (2012). Spontaneous Theory of Mind and Its Absence in Autism Spectrum Disorders. *The Neuroscientist*, 18(2), 108–113. <https://doi.org/10.1177/1073858410397208>
- Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind Eyes: An Absence of Spontaneous Theory of Mind in Asperger Syndrome. *Science*, 325(5942), 883–885. <https://doi.org/10.1126/science.1176170>
- Shanon, C. (1948). *A Mathematical Theory of Communication*. *Bell System Technical Journal*, 27 (4), 623-656.
- Skokauskas, N., & Gallagher, L. (2010). Psychosis, affective disorders and anxiety in autistic spectrum disorder: Prevalence and nosological considerations. *Psychopathology*, 43(1), 8–16.
- Sopory, P., & Dillard, J. P. (2002). The Persuasive Effects of Metaphor: A Meta-Analysis. *Human Communication Research*, 28(3), 382–419. <https://doi.org/10.1111/j.1468-2958.2002.tb00813.x>
- Southgate, V., Senju, A., & Csibra, G. (2007). Action Anticipation Through Attribution of False Belief by 2-Year-Olds. *Psychological Science*, 18(7), 587–592. <https://doi.org/10.1111/j.1467-9280.2007.01944.x>
- Sperber, D. (1994). Understanding verbal understanding. In J. Khalifa (Ed.), *What is intelligence?* (pp. 179–198). Cambridge University Press.
- Sperber, D. (2005). Modularity and relevance. *The Innate Mind: Structure and Contents*, 1, 53–68.
- Sperber, D., & Wilson, D. (1989). *La Pertinence*. Editions de Minuit. [http://www.leseditionsdeminuit.fr/livre-La\\_Pertinence-2269-1-1-0-1.html](http://www.leseditionsdeminuit.fr/livre-La_Pertinence-2269-1-1-0-1.html)

Sperber, D., & Wilson, D. (1996). *Relevance: Communication and Cognition* (2nd ed.). Wiley-Blackwell.

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

Spotorno, N., & Noveck, I. (2014). When is irony effortful? *Journal of Experimental Psychology. General*, *143*(4), 1649–1665. <https://doi.org/10.1037/a0036630>

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., & Krennmayr, T. (2010). Metaphor in usage. *Cognitive Linguistics*, *21*(4). <https://doi.org/10.1515/cogl.2010.024>

Stephan, K. E., & Mathys, C. (2014). Computational approaches to psychiatry. *Current Opinion in Neurobiology*, *25*, 85–92. <https://doi.org/10.1016/j.conb.2013.12.007>

Stephens, G., & Matthews, D. (2014). The communicative infant from 0-18 months: The social-cognitive foundations of pragmatic development. In *Pragmatic development in first language acquisition* (pp. 13–35). John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.10.02ste>

Stoddard, J. A., & Afari, N. (2014). *The Big Book of ACT Metaphors: A practitioner's guide to experiential exercises and metaphors in Acceptance and Commitment Therapy*. new harbinger publications.

Su, Y., & Su, L.-Y. (2015). Interpretation of Logical Words in Mandarin-Speaking Children with Autism Spectrum Disorders: Uncovering Knowledge of Semantics and Pragmatics. *Journal of Autism and Developmental Disorders*, *45*(7), 1938–1950. <https://doi.org/10.1007/s10803-014-2350-0>

Surian, L., Baron-Cohen, S., & van der Lely, H. (1996). Are Children with Autism Deaf to Gricean Maxims? *Cognitive Neuropsychiatry*, *1*(1), 55–72. <https://doi.org/10.1080/135468096396703>

Tannen, D. (1984). *Conversational Style* (Lawrence Erlbaum Associates, Inc.).

Tannen, D. (1991). *You just don't understand: Men and women in conversation*. Ballantine Books.

Thomas, M. S. C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using Developmental Trajectories to Understand Developmental Disorders. *Journal of Speech, Language, and Hearing Research*, *52*(2), 336–358. [https://doi.org/10.1044/1092-4388\(2009/07-0144\)](https://doi.org/10.1044/1092-4388(2009/07-0144))

Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijdsdijk, F. (2016). Heritability of autism

spectrum disorders: A meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry*, 57(5), 585–595.

Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, 10(2), 201–224.

Tonini, E., Bischetti, L., Del Sette, P., Tosi, E., Lecce, S., & Bambini, V. (2023). On the specificity of the relationship between metaphor skills and Theory of Mind in middle childhood: Reviving the Referential Metaphors task. *Cognition*.

Tonini, E., Lecce, S., Del Sette, P., Bianco, F., Canal, P., & Bambini, V. (2022). Efficacy and benefits of the MetaCom training to promote metaphor comprehension in typical development. *First Language*, 42(3), 466–496. <https://doi.org/10.1177/01427237221081201>

Tourangeau, R., & Sternberg, R. J. (1982). Understanding and appreciating metaphors. *Cognition*, 11(3), 203–244.

Trott, S., & Bergen, B. (2019). Individual Differences in Mentalizing Capacity Predict Indirect Request Comprehension. *Discourse Processes*, 56(8), 675–707. <https://doi.org/10.1080/0163853X.2018.1548219>

Trott, S., & Bergen, B. (2020). When Do Comprehenders Mentalize for Pragmatic Inference? *Discourse Processes*, 57(10), 900–920. <https://doi.org/10.1080/0163853X.2020.1822709>

Valle, A., Massaro, D., Castelli, I., & Marchetti, A. (2015). Theory of Mind Development in Adolescence and Early Adulthood: The Growing Complexity of Recursive Thinking Ability. *Europe's Journal of Psychology*, 11(1), 112–124. <https://doi.org/10.5964/ejop.v11i1.829>

van Arkel, J., Woensdregt, M. S., Dingemanse, M., & Blokpoel, M. (2020). A simple repair mechanism can alleviate computational demands of pragmatic reasoning: Simulations and complexity analysis. <https://www.aclweb.org/anthology/2020.conll-1.14/>. 24th Conference on Computational Natural Language Learning. <https://repository.ubn.ru.nl/handle/2066/226703>

Van Herwegen, J., Dimitriou, D., & Rundblad, G. (2013). Development of novel metaphor and metonymy comprehension in typically developing children and Williams syndrome. *Research in Developmental Disabilities*, 34(4), 1300–1311. <https://doi.org/10.1016/j.ridd.2013.01.017>

Van Herwegen, J., & Rundblad, G. (2018). A Cross-Sectional and Longitudinal Study of Novel Metaphor and Metonymy Comprehension in Children, Adolescents, and Adults With Autism Spectrum Disorder. *Frontiers in Psychology*, 9, 945.

<https://doi.org/10.3389/fpsyg.2018.00945>

Van Steensel, F. J., Bögels, S. M., & Perrin, S. (2011). Anxiety disorders in children and adolescents with autistic spectrum disorders: A meta-analysis. *Clinical Child and Family Psychology Review*, *14*, 302–317.

Vivanti, G. (2020). Ask the Editor: What is the Most Appropriate Way to Talk About Individuals with a Diagnosis of Autism? *Journal of Autism and Developmental Disorders*, *50*(2), 691–693. <https://doi.org/10.1007/s10803-019-04280-x>

Vulchanova, M., Saldaña, D., Chahboun, S., & Vulchanov, V. (2015). Figurative language processing in atypical populations: The ASD perspective. *Frontiers in Human Neuroscience*, *9*, 24. <https://doi.org/10.3389/fnhum.2015.00024>

Vulchanova, M., & Vulchanov, V. (2022). Rethinking Figurative Language in Autism: What Evidence Can We Use for Interventions? *Frontiers in Communication*, *7*, 910850. <https://doi.org/10.3389/fcomm.2022.910850>

Watkins, L., O'Reilly, M., Kuhn, M., Gevarter, C., Lancioni, G. E., Sigafos, J., & Lang, R. (2015). A Review of Peer-Mediated Social Interaction Interventions for Students with Autism in Inclusive Settings. *Journal of Autism and Developmental Disorders*, *45*(4), 1070–1083. <https://doi.org/10.1007/s10803-014-2264-x>

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388. <https://doi.org/10.1016/j.tics.2010.05.006>

Wearing, C. (2010). Autism, Metaphor and Relevance Theory. *Mind & Language*, *25*(2), 196–216. <https://doi.org/10.1111/j.1468-0017.2009.01386.x>

Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. *Frontiers in Human Neuroscience*, *8*. <https://doi.org/10.3389/fnhum.2014.00583>

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, *72*(3), 655–684. <https://doi.org/10.1111/1467-8624.00304>

Whalen, J. M., Pexman, P. M., & Gill, A. J. (2009). “Should Be Fun—Not!”: Incidence and Marking of Nonliteral Language in E-Mail. *Journal of Language and Social Psychology*, *28*(3), 263–280. <https://doi.org/10.1177/0261927X09335253>

Whitehouse, A. J. O., Watt, H. J., Line, E. A., & Bishop, D. V. M. (2009). Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism. *International Journal of Language & Communication Disorders*, *44*(4), 511–528.

<https://doi.org/10.1080/13682820802708098>

Whyte, E. M., & Nelson, K. E. (2015). Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *Journal of Communication Disorders*, *54*, 2–14. <https://doi.org/10.1016/j.jcomdis.2015.01.001>

Willinger, U., Deckert, M., Schmöger, M., Schaunig-Busch, I., Formann, A. K., & Auff, E. (2019). Developmental Steps in Metaphorical Language Abilities: The Influence of Age, Gender, Cognitive Flexibility, Information Processing Speed, and Analogical Reasoning. *Language and Speech*, *62*(2), 207–228. <https://doi.org/10.1177/0023830917746552>

Wilson, A. C. (2022). Development and validation of the conversation questionnaire: A psychometric measure of communication challenges generated from the self-reports of autistic people. *Autism & Developmental Language Impairments*, *7*, 239694152211232. <https://doi.org/10.1177/23969415221123286>

Wilson, A. C., & Bishop, D. V. (2019). “If you catch my drift...”: Ability to infer implied meaning is distinct from vocabulary and grammar skills. *Wellcome Open Research*, *4*, 68.

Wilson, A. C., & Bishop, D. V. M. (2021). “Second guessing yourself all the time about what they really mean...”: Cognitive differences between autistic and non-autistic adults in understanding implied meaning. *Autism Research*, *14*(1), 93–101. <https://doi.org/10.1002/aur.2345>

Wilson, D., & Sperber, D. (2006). Relevance Theory. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 606–632). Blackwell Publishing Ltd. <https://doi.org/10.1002/9780470756959.ch27>

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition*, *13*, 103–128.

Wing, L. (1997). The autistic spectrum. *The Lancet*, *350*(9093), 1761–1766.

Winner, E. (1997). *The point of words: Children’s understanding of metaphor and irony*. Harvard University Press.

Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, *12*, 289–297. <https://doi.org/10.1037/0012-1649.12.4.289>

Wolters, N., Knoors, H., Cillessen, A. H. N., & Verhoeven, L. (2014). Behavioral, Personality, and Communicative Predictors of Acceptance and Popularity in Early Adolescence. *The Journal of Early Adolescence*, *34*(5), 585–605.

<https://doi.org/10.1177/0272431613510403>

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*, 22(2), 145–160.

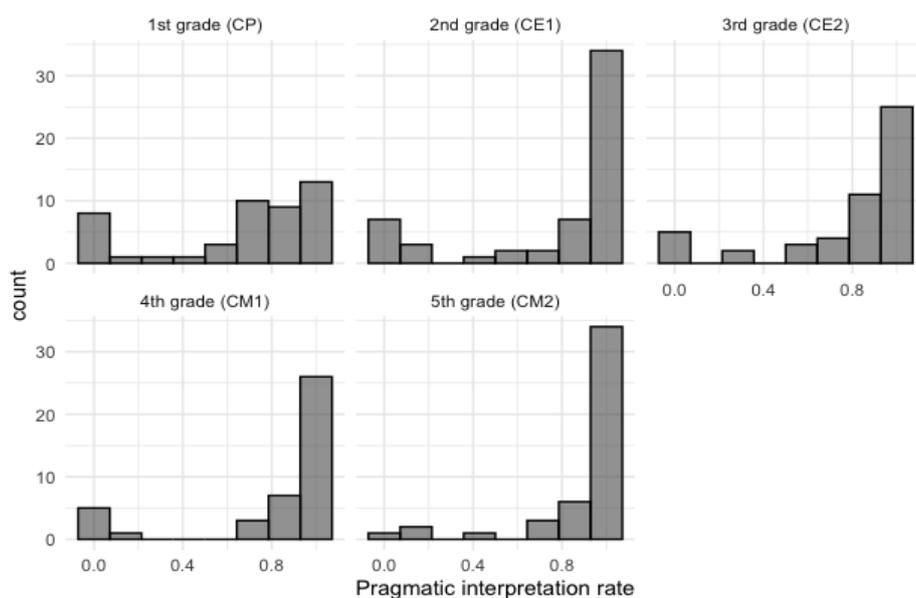
World Health Organization. (2019). *International Classification of Diseases, Eleventh Revision (ICD-11)* (11th ed.). <https://icd.who.int/browse11>

Zalla, T., Amsellem, F., Chaste, P., Ervas, F., Leboyer, M., & Champagne-Lavau, M. (2014). Individuals with autism spectrum disorders do not use social stereotypes in irony comprehension. *PloS One*, 9(4), e95568. <https://doi.org/10.1371/journal.pone.0095568>

Zhao, M., Liu, X., Dai, X., Dong, S., & Han, Z. (2021). Scalar implicature is not a default process: An ERP study of the scalar implicature processing under the effect of focus factor. *Brain Research*, 147499. <https://doi.org/10.1016/j.brainres.2021.147499>

# ANNEXES

## Annexe A • Matériel supplémentaire pour l'étude 1



**Figure 1:** Pragmatic interpretations rates distributions per grade

**Table 1:** post-hoc contrasts of response-types comparisons at different ages ( $\beta$  = estimates (SE); significant (\*) or marginally significant (~) comparisons in bold)

Age	6 years old	7 years old	8 years old	9 years old	10 years old	11 years old
control - some <sub>pragma</sub>	<b><math>\beta = -0.13</math> (0.05), <math>p = .04^*</math></b>	<b><math>\beta = -0.10</math> (0.03), <math>p = .03^*</math></b>	<b><math>\beta = -0.09</math> (0.04), <math>p = .07\sim</math></b>	<b><math>\beta = -0.09</math> (0.04), <math>p = .04^*</math></b>	<b><math>\beta = -0.11</math> (0.03), <math>p = 0.01^*</math></b>	<b><math>\beta = -0.15</math> (0.05), <math>p = .01^*</math></b>
control - some <sub>sem</sub>	$\beta = 0.13$ (0.07), $p = .15$	$\beta = 0.03$ (0.04), $p = .78$	$\beta = -0.05$ (0.04), $p = .56$	<b><math>\beta = -0.09</math> (0.04), <math>p = .08\sim</math></b>	<b><math>\beta = -0.12</math> (0.05), <math>p = .07\sim</math></b>	$\beta = -0.11$ (0.10), $p = .48$
some <sub>sem</sub> - some <sub>pragma</sub>	<b><math>\beta = -0.25</math> (0.08), <math>p = .003^*</math></b>	<b><math>\beta = -0.13</math> (0.05), <math>p = .04^*</math></b>	$\beta = -0.04$ (0.05), $p = .74$	$\beta = 0.003$ (0.05), $p = .99$	$\beta = 0.004$ (0.06), $p = .99$	$\beta = -0.04$ (0.11), $p = .93$

## **Annexe B** • Évaluer la théorie de l'esprit chez l'enfant (étude annexe)

Article *en révision* pour *Child Psychiatry and Human Development*: Assessing theory of mind in children: a tablet-based adaptation of a classic picture sequencing task

Nicolas PETIT, Ira NOVECK, Matias BALTAZAR, Jérôme PRADO

---

### **Abstract**

Correctly assessing children's theory of mind (TOM) is essential in clinical practice. Yet, most tasks heavily rely on language, which constitutes an obstacle for several populations. Langdon and Coltheart's (1999) Picture Sequencing Task (PST), developed for research, avoids this limitation through a minimally-verbal procedure. We thus developed a tablet adaptation of this task for individual application, engaging children's motivation and allowing response times collection. To assess this tablet-PST, we first tested a large sample of neurotypical children (6-11 years-old, N = 248), which confirmed the task's structural and content validity, and permitted the construction of three standardized clinical indices. In a second experiment, we applied those to previously diagnosed autistic children (N = 23), who were expected to show atypical TOM performances. Children's outcomes were consistent with what was hypothesized and confirmed the task's external validity and moderate clinical sensitivity. The tablet-PST thus appears as a suitable tool, providing detailed profiles to inform clinical decisions.

---

*Theory of mind* (TOM) refers to the ability to attribute mental states in order to explain and predict people's behaviors (Premack & Woodruff, 1978). TOM is thought to be essential for human social interactions, grounding communication, self-representation, and moral reasoning (Waytz et al., 2010). Although TOM emerges in the first months of development (Frith & Frith, 2003), studies have found that TOM skills keep maturing up to late adolescence (e.g.,

Bosco et al., 2014). In the experimental literature, TOM is classically operationalized and assessed by measuring the ability to understand false beliefs and to distinguish such beliefs from one's own perception of the world (Dennett, 1978). In seminal false belief task, Wimmer & Perner (1983) presented children with situations where a character wrongly believed that a chocolate was in a given location, while children knew that the chocolate was in

elsewhere. This revealed an important milestone around the age of 4, at which children can be expected to understand false beliefs (Wellman et al., 2001). This paradigmatic situation was also used to assess the representation of second order mental states (Mary *believes* that John *thinks* that X), there are arguably more complicated to process and that are mastered only later in development, around the age of 7 (Perner & Wimmer, 1985), and TOM then continues to mature during school-age (e.g., O'Hare et al., 2009). Beyond the understanding of false beliefs, the operationalization of TOM in assessment tools has widened to encompass the comprehension of different types of mental states, such as hidden intentions or affective states, and a variety of tasks exists for pre-school-age children (Beaudoin et al., 2020). Tasks targeting school-age, however, appeared to display relatively low psychometric properties in neurotypical children (Hayward & Homer, 2017).

Critically, TOM abilities are also characterized by interindividual differences that relate to children's social and cognitive skills. For example, early TOM skills can predict later social preference, friendship quality or academic achievement (Blair & Razza, 2007; Dunn et al., 2002; Fink et al., 2014). TOM impairments also affect quality of life in several clinical populations, such as schizophrenic patients (Maat et al., 2012), individuals with traumatic brain injuries (Bivona et al., 2015), children with developmental language disorders (Botting & Conti-Ramsden, 2008) and children with autism spectrum disorder (from now on: *autism*). In clinical practice, it is therefore of obvious importance to correctly assess TOM skills

in order to detect a patient's difficulties and guide interventions.

Autism is exemplary of a case that involves TOM difficulties, as these impairments have been postulated to be central or primary to this condition (Baron-Cohen, 1997). This led to a large body of research that eventually rephrased it in terms of *differences* instead of a *deficit*, and that questioned both the primacy and the universality of this marker while confirming the overall association between autism and TOM divergences (see Fletcher-Watson & Happé, 2019, but also Marocchini, 2023). For example, Senju (2012) proposed that autistic individuals<sup>23</sup> could be able to pass standard false belief tasks when asked to, but that they would struggle to *spontaneously* use TOM to predict others' behavior. This reveals that identifying autistic-neurotypical differences could require subtle experimental designs (e.g., as opposed to simply observing mean accuracy on TOM tasks, see Baltazar et al., 2021). In line with this, Livingston & Happé (2017) argued that autistic individuals might achieve the same level of performance as neurotypical participants in some highly-controlled lab-environments, but that they could rely on atypical processes that might be unusually effortful for them. This explains why some researchers employ response times (RT) to better characterize TOM mechanisms (e.g., Behrmann et al., 2006; Livingston et al., 2019). For example, Kaland et al. (2007), who compared autistic to neurotypical adolescents, showed that mental state inferences generally take longer to process than physical state inferences, but that this slowdown was more pronounced in autism.

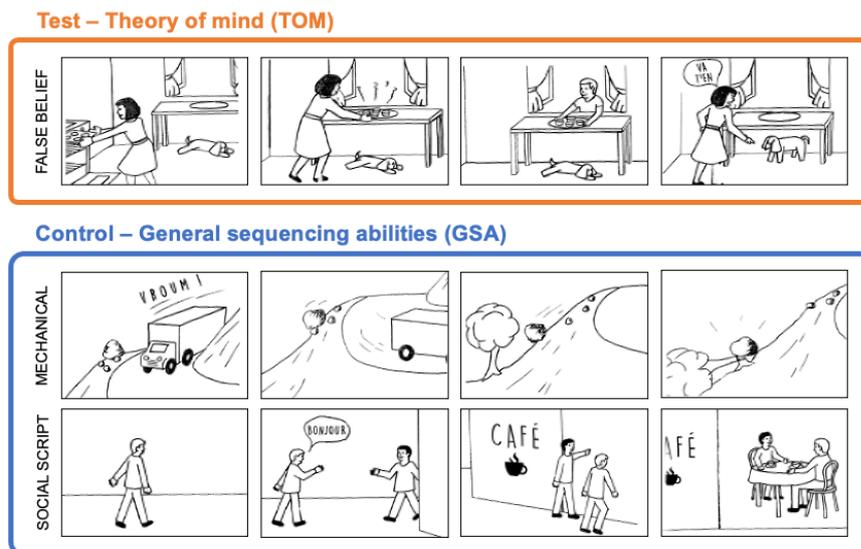
---

<sup>23</sup> Following Botha et al. (2021), we use in this paper identity-first language to refer to autistic individuals, judged less offending and stigmatizing than person-first language (“person

with autism”) by the autistic community (Bury et al., 2020). Note that the adjective “autistic” refers here to the whole autistic spectrum.

It is also important to note that most classic TOM tasks that are used in clinical practice largely rely on verbal material. It is the case of the standard false belief task (Wimmer & Perner, 1983), which uses verbal stories and verbal questions. However, it is also the case of other tasks that have been developed to assess more complex aspects of TOM (i.e., developing later than 1<sup>st</sup> order false beliefs, see Hayward & Homer [2017]). For example, the Strange Stories test (Happé, 1994), the Faux-pas test (Baron-Cohen et al., 1999) or the Reading the Mind in the Eyes Test (Baron-Cohen et al., 2001) are among the mostly used, but all either use verbally presented stories or rely on a verbal response format. This is an important confound, because language and TOM are known to be interrelated skills across development (de Villiers, 2007; Milligan et al., 2007; Pyers & Senghas, 2009). Such tasks are thus difficult to use in individuals with language disorders because failures can be attributed to either poor TOM or verbal difficulties. This is particularly relevant for autism, which is often found associated with comorbid conditions, especially language disorders (Levy et al., 2010).

Interestingly, TOM needs not be assessed using verbal tasks. For example, Langdon & Coltheart (1999) developed the Picture Sequencing Task (*PST* from now on) precisely to measure TOM without using complex verbal material. In this task, participants are asked to order four picture-cards such that it forms a coherent story. Each picture contains no linguistic material, or very basic 1-word or 2-words utterances (e.g., “Chocolate”, “Hi”, “Vroom”, “Go away”). Critically, these stories can involve characters acting upon false beliefs and, as such, require TOM. Two other types of stories are used as controls (see Figure B-1): mechanical stories (involving physical causalities) and social-scripts (involving everyday social routines). The task includes four sequences of each type, and each of the sequence is scored between 0 and 6. The *PST* has been used successfully to isolate TOM skills in a variety of typical or clinical populations (Langdon & Coltheart, 1999; Payne et al., 2016; Porter et al., 2008; Rajkumar et al., 2008; Van Rheenen & Rossell, 2013). For instance, in Langdon and Coltheart’ (1999) study, the *PST* revealed that healthy adults with high schizotypal traits differed from low-schizotypal controls in the TOM condition



**Figure B-1** : *PST* stories example, involving false beliefs to assess theory of mind (above), or mechanical causalities and social scripts as control (below)

only, but not in the control conditions. In neurotypical children, Rajkumar et al. (2008) showed that this task could produce a TOM index that was independent from general intelligence but correlated with another measure of social intelligence (Rajkumar et al., 2008). To our knowledge, however, no measurement of the psychometric characteristics of this task has ever been reported.

Moreover, to date, the PST has only been used in group designs, for example to compare average TOM skills in groups of patients versus normal controls. The present study reports a computerized version of the PST (prepared for tablet computers) to be used as a clinical tool in order to assess the TOM skills of a single child. Compared to paper-and-pencil solutions, assessments performed on computer tablets tend to be preferred by children (Piatt et al., 2016) and favor engagement and attention. This is especially the case in autistic children (Lane & Radesky, 2019), who may benefit from such interaction-free and predictable environments. Tablet-based tests also allow for perfectly standardized instruction, execution, feedback and scoring procedures, reducing the risk of experimenter bias or errors, while allowing for reliable group-testing (Bignardi et al., 2021). Moreover, tablets offer an easy and reliable solution to record response times, which would provide relevant complementary information (besides just accuracy)<sup>24</sup>.

In this study, we aimed to evaluate whether the tablet-PST could be used to assess individual children's theory of mind skills. The study also aimed to provide developmental normative data for the task. More specifically, we made four

hypotheses regarding the characteristics of the tablet-PST. First, it should show a good face validity as evidenced by a difference in difficulty between sequences that involve TOM and those that do not. Second, it should show a good structural validity, as evidenced by confirming a three-dimensions structure (along with the three types of sequences that it includes) and the involvement of a shared ability to sequence pictures in general. Third, it should be sensitive to both (i) individual differences between children and (ii) group difference between neurotypical and autistic children. Specifically, the tablet-PST should discriminate these two groups, although not perfectly since TOM is not considered as divergent rather than absent in autism. Finally, consistent with prior research that using the PST (Rajkumar et al., 2008), the tablet-PST should also demonstrate a good external validity within the group of autistic children by providing results that are consistent with a gold-standard TOM task.

## Experiment 1

In order to meet our goal, we begin by testing a large sample of neurotypical children with the tablet-PST, and assess its psychometric properties in terms of structural and content validity, and sensitivity to interindividual differences. School-age is arguably as an appropriate developmental period, during which TOM skills are maturing and can be the target of intervention (see Westby & Robinson, 2014; or Rajkumar et al., 2008, for data using the PST). It is also an age that is still critical for the diagnosis of neurodevelopmental disorders (e.g., see

---

<sup>24</sup> Interestingly, response times measures were included in Langdon & Coltheart (1999)'s original study but were later abandoned, after failing to reveal group difference between high

vs. low schizotypal healthy adults. Having seen that response times could reveal meaningful effects in autism, we decided to include this measure on the tablet-PST.

Van 'T Hof et al., 2021, for autism), but also where only few assessment tools display good psychometric properties (Hayward & Homer, 2017). Provided that such measures are satisfying, this will also allow for the construction of normative data which could be used to derive standardized scores for individual applications.

## Methods

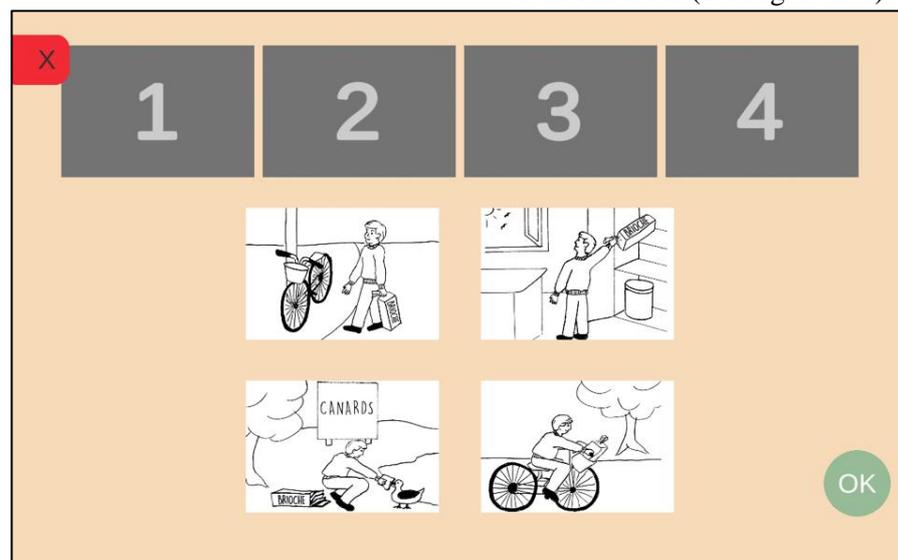
### Participants

We recruited 248 children in primary schools for this experiment (mean age = 8.3 years old, min = 6.0, max = 11.0, 51% female). All participants were native French speakers. Fifty-two participants were in 1<sup>st</sup> grade, 58 in 2<sup>nd</sup> grade, 50 in 3<sup>rd</sup> grade, 42 in 4<sup>th</sup> grade and 46 in 5<sup>th</sup> grade. Parental information and consent were collected before testing. All parents declared that their child presented no learning or neurodevelopmental disorder (i.e., autism, attention deficit, or language-related disorder) and no sensory (visual or auditory) or motor disorders that would interfere with their ability to use a tablet. Parents also provided information regarding their socio-economic

environment, through (1) the Family Affluence Scale (FAS, Currie et al., 2008), which is a good indicator of family wealth across countries, and (2) their education levels, which were coded on a scale from 0 (no diploma) to 7 (PhD). The FAS and the parental education level were standardized on a scale ranging from 0 to 100 (representing the range of minimum to maximum possible values) and averaged to provide a socio-economic status composite (SES; mean = 58.4; SD = 16.1). This study was part of a larger project on pragmatic inference skills across development, which received ethical approval from the local IRB (Comité de Protection des Personnes Sud-Est I, ID RCB 2019-A01721-56).

### Material

As the original task, the tablet-PST includes four trials of each type (mechanical, social scripts, false beliefs) as well as two training sequences. Each of these will be referred to as an *item*. The minimal verbal content of the original pictures (as depicted in Figure B-1) was translated into French. The task was implemented in an application designed for tablet computers. At the beginning of each trial, four pictures were presented in the lower part of the screen in a random order (see Figure B-2) while four



**Figure B-2:** Task set up for a practice item as it appears on the tablets of the participants

empty slots were located in the upper part of the screen. Children had to select and move the four pictures to the appropriate slots so that the sequence was in order and described a coherent story. They were allowed to modify the sequence as often as they wanted after moving the pictures if they changed their mind. When satisfied, they validated the sequence by pressing *OK* on the bottom right side of the screen. In accordance with the original procedure (Langdon & Coltheart, 1999), accuracy on each item ranged from 0 to 6 based on the appropriate position of each picture (2 points for the first and the last picture, 1 point for the 2<sup>nd</sup> and the 3<sup>rd</sup>). Accuracy as well as response time (RT, i.e., time from the beginning of the trial to validation) was recorded by the device at the end of each trial. Trials were pseudo-randomized in a unique trial list. Feedback was automatically provided by the device for the 2 training sequences (a green screen in case of success, a grey screen in case of failure), which children had to perform successfully to move on to the rest of the task (these were repeated as many times as necessary).

### Procedure

Assessment took place in children's usual classrooms, i.e., in a collective setting, with 15 to 25 children completing the task simultaneously. Each child was provided with a tablet. Automatic audio instructions were included in the task, asking children to move the four pictures to the appropriate slots so that the sequence was in order and described a coherent story, for each sequence. The task began with the two practice items, during which two experimenters remained available and could help children to correctly sequence pictures through verbal guidance. Then the

task automatically continued with task items, during which no help was provided to children. Children performed the tasks at their own pace and were asked to remain quiet. During task completion, two experimenters remained present to make sure that children would not copy each other. The task, which lasted about 10 minutes, was performed after other experimental tasks measuring children's pragmatic skills (not reported here) as part of a larger assessment that took about 20 minutes.

### Analysis

All analyses were conducted in R (R Core Team, 2022). These involved three steps. First, we assessed the internal consistency and the factorial structure of the task. Internal consistency was assessed with Cronbach's alpha. The impact of each item on this value was also systematically measured. Factorial structure was assessed with confirmatory factor analysis (CFA) using the *lavaan* package (Rosseel, 2012). Specifically, we compared a unifactorial structure to a theoretical three-dimensions structure.

Second, we used mixed effect models on single trials to analyze accuracy and response times at the group level. In these analyses, we were not interested in the differences among the control items (i.e., between the mechanical and social scripts), but in using them as controls for false-belief items. We thus followed the rationale of Langdon & Coltheart, 1999 (see also Rajkumar et al., 2008 for children) and contrasted TOM items to non-TOM items as a binary variable (*item-type*).<sup>25</sup> The model with accuracy as dependent variable included item-type, age, gender and SES as fixed effects, as well as the interactions of

---

<sup>25</sup> To make sure that grouping together social scripts and mechanical did not hide important effects, we also run the analyses contrasting

TOM, mechanical and social scripts, and report the results in supplementary materials.

item-type with the latter three. In addition to these factors, the model with response times as dependent variable also included accuracy on each trial and its interaction with item-type as an additional fixed factor. All models included random intercepts for participants and items. Age was included as a continuous predictor (in years with decimals), considering both a linear and a quadratic term (to capture potential non-linear effects). The response time analysis was run on log-transformed values to correct for skewness. Outliers were removed, first roughly at a group level by discarding trials with a response time shorter than 5 seconds or longer than 80 seconds, and secondly at a participant-specific level by excluding trials with a response time greater and lesser than 2.5 SD from each individual's mean RT. Models were fit with the *lme4* package (Bates et al., 2015), sum contrasts were used for all factors, and continuous predictors were centered on the mean. Effects were assessed with Satterthwaite's degrees of freedom from the *lmerTest* package (Kuznetsova et al., 2017), and post-hoc contrasts were computed with the *emmeans* package (Lenth, 2022).

Third, for the sake of preparing the clinical application, we calculated summary measures for each participant and used it to fit linear regressions. These regressions were meant to allow for the comparison of an individual's actual performance to their predicted performance. The advantage of using regressions to calculate discrepancy between a single participant and a reference group is that, instead of considering arbitrarily-defined control sub-groups to provide normative data (for example age-groups or SES-groups), the whole sample size and its variability is exploited to improve the statistical power of standardized scores calculation (Crawford & Garthwaite, 2006). Rather than specifying *a priori* the predictors to be

included in those models, we used the factors that were shown to be significant at group-level (step 2); that is, for instance, we specified age as predictor if and only if it was shown to be linked to performance. To anticipate, based on group-level results, we selected age and SES as predictors of children's average accuracy on non-TOM items (as a General Sequencing Abilities index, GSA), we used this GSA and both a linear and quadratic term for age to predict children average accuracy on TOM items, and we used the mean response time to non-TOM items and the average accuracy on TOM items to predict the average response time to TOM-items.

## Results

### Task validation

Across all items, the overall Cronbach alpha was .75, suggesting acceptable internal consistency. Importantly, removing each item decreased this value (for details, see the supplementary materials, Table S.1.1). The CFA revealed that the three-factor structure (TOM, Mechanical items, Social-scripts) provided a good fit to the data ( $CFI = .93$ ,  $TLI = .90$ ,  $RMSEA = .047$ ), which was significantly better than a unidimensional model ( $X^2(3) = 17.8$ ,  $p < .001$ ). The three factors covaried (standardized estimates varied from .63 to .90, all  $ps < .001$ , see Figure S1.1 in supplementary materials for detailed estimates).

### Group analysis

The linear mixed effect model with accuracy as dependent variable (Table B-1) revealed a main effect of item-type, indicating that participants were less accurate on TOM items than the control (non-TOM) items. However, item-type

**Table B-1:** Estimated parameters of the model explaining accuracy

Parameter	Coefficient	Standard error	<i>p</i>
(Intercept)	3.14	0.26	<.001
<b>item-type</b>	<b>0.95</b>	<b>0.39</b>	<b>.009</b>
gender	0.05	0.11	.659
<b>age</b>	<b>22.06</b>	<b>3.09</b>	<b>&lt; .001</b>
age <sup>2</sup>	-5.83	3.09	.059
<b>SES</b>	<b>0.02</b>	<b>0.00</b>	<b>&lt; .001</b>
item-type * gender	0.16	0.14	.237
item-type * age	-4.95	3.71	.182
<b>item-type * age<sup>2</sup></b>	<b>8.59</b>	<b>3.70</b>	<b>.020</b>
item-type * SES	0.00	0.00	.445

*age<sup>2</sup> = quadratic term for age*

interacted with age: while accuracy on non-TOM items increased linearly with age, accuracy on TOM items reached a plateau at around 9-10 years old (see Figure B-3). There was also a main effect of SES, indicating that accuracy generally increased with higher SES. No other main effect or interaction was significant.

The linear mixed model with response times as a dependent variable (see Table B-2) revealed a significant interaction between item-type and accuracy (see Figure S1.2 in supplementary materials). While response times did not depend on accuracy for control items ( $\beta = -0.01$ ,  $SE = 0.01$ ), it increased with accuracy on TOM items ( $\beta = 0.03$ ,  $SE = 0.01$ ). There was also an interaction between item-type and age. Responses become faster overall with age,

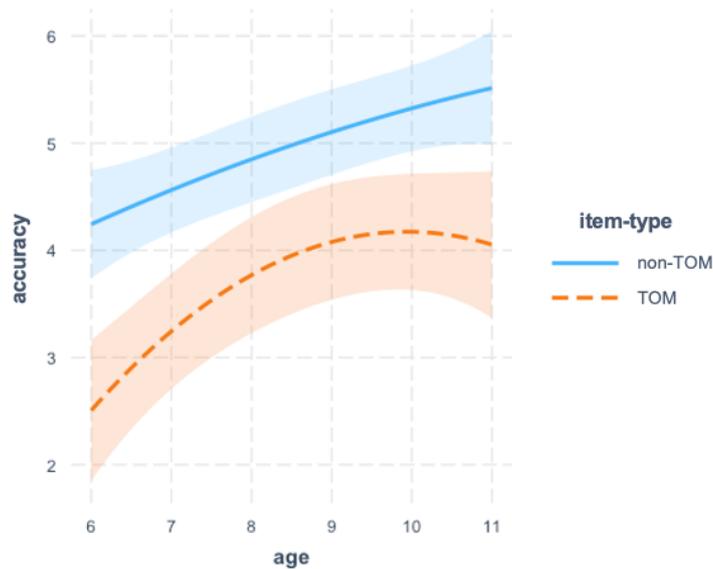
reaching a plateau at roughly 9 years of age. TOM items were associated with longer response times when compared to non-TOM items and this difference increased with age.

To make sure that the social script/mechanical story grouping did not hide important effects, we exploratorily reran these two models with 3 item-types instead, which revealed the very same patterns (see in Supplementary materials the detailed models' outputs in Tables S1.2 and S1.3, and the predictions in figures S1.3 and S1.4).

**Table B-2:** Estimated parameters of the model explaining log-transformed response times

Parameter	Coefficient	Standard error	<i>p</i>
(Intercept)	9.81	0.08	< .001
item-type	0.07	0.11	.553
<b>accuracy</b>	<b>0.01</b>	<b>0.00</b>	<b>.004</b>
SES	0.00	0.00	.126
<b>age</b>	<b>-4.32</b>	<b>0.99</b>	<b>&lt; .001</b>
<b>age<sup>2</sup></b>	<b>2.55</b>	<b>0.98</b>	<b>.010</b>
gender	-0.07	0.04	.056
<b>item-type * score</b>	<b>0.03</b>	<b>0.01</b>	<b>&lt; .001</b>
item-type * SES	0.00	0.00	.175
<b>item-type * age</b>	<b>2.99</b>	<b>0.79</b>	<b>&lt; .001</b>
item-type * age <sup>2</sup>	0.59	0.77	.446
item-type * gender	-0.01	0.03	.774

*age<sup>2</sup> = quadratic term for age*



**Figure B-3:** model's predicted accuracy as a function of age and item-type (shaded areas represent 95% confidence intervals)

### Individual indices

Summary measures were then calculated for each individual. Specifically, we calculated a TOM-accuracy score by averaging accuracy across TOM items, as well as a General Sequencing Abilities (GSA) score by averaging accuracy on control items (following Rajkumar et al., 2008). Both indices showed satisfying distributions, avoiding floor or ceiling effects across age groups, although the GSA approached ceiling by 5<sup>th</sup> grade (see Figure S1.5 in the supplementary materials). Log-transformed response times were also averaged for these two item-types (TOM and non-TOM).

Based on the effects observed at the group-level (see above), three linear models were fitted to allow for the calculation of individual discrepancy indices (which will be useful for Experiment 2 below). First, we calculated a model explaining each child's GSA based on their age ( $\beta = 0.26$ ,  $SE = 0.04$ ,  $p < .001$ ) and SES ( $\beta = 0.02$ ,  $SE = 0.004$ ,  $p < .001$ ). This provides a control index of the extent to which a child is able to sequence pictures in general. Second, we calculated a model explaining each child's

TOM-Accuracy based on their GSA ( $\beta = 0.52$ ,  $SE = 0.07$ ,  $p < .001$ ) and their age with a linear ( $\beta = 1.94$ ,  $SE = 0.69$ ,  $p < .001$ ) and a quadratic term ( $\beta -0.0007$ ,  $SE = 0.00028$ ,  $p = .01$ ). This allows one to characterize variations of accuracy on TOM items after regressing out children's age and ability to sequence pictures in general. Third, we calculated a model explaining each child's TOM-Response times, based on their accuracy (accounting for speed-accuracy tradeoff,  $\beta = 0.057$ ,  $SE = 0.01$ ,  $p < .001$ ) and their own baseline RT on control items ( $\beta = 0.90$ ,  $SE = 0.05$ ,  $p < .001$ ). This model allows one to estimate the effort specifically associated with TOM items.

### Discussion

In this experiment, we tested a tablet implementation of the PST on 248 neurotypical school-aged children. The confirmatory factor analysis provided evidence in support of the structural validity of the tablet-PST, with 3-dimensions opposing items involving mechanical causalities, social scripts and false beliefs. The non-negligible shared variance

between the three factors (also visible in the Cronbach alpha demonstrating internal consistency) confirms the appropriateness of averaging control items together in order to be compared to TOM items, following previous uses of the PST (Langdon & Coltheart, 1999; Rajkumar et al., 2008). Even though we did not directly compare the results of the original paper-and-pencil PST to those of the tablet-PST in our sample, our results are in line with previous studies which used the original task with healthy participants, and which typically observed that TOM items were harder than mechanical and social-scripts items. This is visible with respect to both accuracy (Payne et al., 2016; Porter et al., 2008; Rajkumar et al., 2008; Van Rheenen & Rossell, 2013) and response times (Langdon & Coltheart, 1999). It is interesting to note that children in our sample performed relatively well overall and better than children from a previous study by Rajkumar et al. (2008), who also tested a group of neurotypical children (between the ages of 8 and 11). Although it might be argued that the difference comes from the tablet implementation of our task, this is not likely because tablet-devices have provided reliable results compared to paper-and-pencil tests elsewhere (Piatt et al., 2016). Another possibility, which could be investigated in future studies, is that the difference stems from cultural or educational differences between our setting (France) and theirs (Southern India). Critically, however, this difference is orthogonal to the contrast observed between TOM and control items. That is, our results remain in line with the TOM literature, which shows that TOM continues to develop during school age (e.g., O'Hare et al., 2009) and that intelligence in general can account for some, but not all variance on TOM performance (Wellman, 2018). This explains why TOM and the control items are linked but follow slightly different developmental patterns.

Langdon & Coltheart (1999), followed by Rajkumar et al. (2008), previously used the PST to calculate a TOM composite by subtracting accuracy in control items to accuracy in TOM items. We followed this general approach for our clinical index, except that we determined the coefficient linking control to TOM accuracy on the basis of the group-level data. Moreover, we added age as a supplementary predictor, since we observed that both item-types followed a different developmental pattern. By doing so, our indices are based on theoretical consideration, but are also data-driven, i.e., they are based on the effects which emerged from the reference data. For example, the control index targeting how easy it is for a child to sequence pictures in general (GSA) was based on age and SES, since both factors played an important role at a group-level, regardless of item-type. Conversely, SES was not included in the TOM-Accuracy index because no interaction was observed between item-type and SES. Including GSA as a predictor of TOM accuracy should thus be enough to account for that variation. For the same reason, gender was not taken into consideration to predict a child's performance. Even though some studies suggested a TOM advantage for girls (Baron-Cohen, 2002), this was not evident in our data, in line with Charman et al. (2002)'s large study that observed this slight advantage, but only in early pre-school development.

Overall, this experiment provided outcomes that were consistent with prior findings in the literature, demonstrating the face validity of the tablet-PST. This also constituted normative data against which individual data could be compared. A next important step would be to find support for the task's utility by applying the clinical indices to a sample of children with a condition known to experience difficulties

in theory of mind, such as autism. This is what we turn to in the next experiment.

## Experiment 2

In this second experiment, we turned to a clinical population to pursue the validation of the tablet-PST for clinical use. Autistic children with normal language and intellectual skills appeared as an appropriate clinical condition for that purpose since, as described in introduction, they are expected to display TOM divergence compared to neurotypical children, but not an absence of TOM; the task should capture this by discriminating these two sub-groups, although not perfectly. Moreover, in this sample we expect a correlation between the scores of the critical condition of the tablet-PST and those of a paradigmatic false-belief story task. In contrast, a much weaker (or even lack of) correlation should be observed with a parental report of Attention Deficit / Hyperactivity Disorder (ADHD) symptoms, which is a distinct psychological construct. This would demonstrate the external validity of the tablet-PST, both in a convergent and divergent way.

## Methods

### Participants

We recruited a sample of 23 autistic children (3 girls), mainly through the department of neurodevelopmental disorders at the *Centre Hospitalier Le Vinatier (Lyon, France)*. All children met the DSM-5 criteria for autism spectrum disorder and received their diagnosis from an experienced child psychiatrist, with no associated intellectual disability, language disorder or attention deficit disorder. Children's assessment results regarding

their autism diagnosis (through the SCQ and/or ADOS-2), as well as their evaluation of intelligence were extracted from their medical records (see Table S2.1 in the supplementary materials), which confirmed the diagnosis. Regarding intellectual abilities, recent measures were not always available (data were not available for three children) and could not be averaged since they were collected with different instruments and gathered in different composites. However, a careful look at all individual results revealed that all children fall in the range of what is considered normal to normal-superior intelligence (all scores are in the 70-130 range, with most scores in the 100-130 range). All participants attended regular schools. Children were between 6;7 and 10;8 years old (mean = 8.7 years old,  $SD = 1.36$ ). SES measures were collected as they were in Experiment 1 and were found to be comparable to those of the control group (mean = 55.2,  $SD = 17.0$ ,  $t(25.8) = 1.40$ ,  $p = .17$ ). Parental information and consent were collected before testing. This experiment was part of the same project than the first experiment, that received ethical approval from the local IRB (Comité de Protection des Personnes Sud-Est I, ID RCB 2019-A01721-56).

### Materials

Children were tested with the tablet-PST, as in Experiment 1. To assess convergent validity, children were also tested with another standard 1<sup>st</sup> and 2<sup>nd</sup> order false belief task. This task, taken from the French battery EVALEO 6-15 (Launay et al., 2018), relies on verbal stories with picture support and verbal open questions targeting characters' (false) beliefs. Stories from this task are directly inspired by classical false belief stories from Wimmer & Perner (1983) or Perner & Wimmer (1985). Five stories are presented, providing a total raw score ranging from 0 to 7. It showed an

acceptable internal consistency in the large sample of children of the validation study ( $\alpha = .78$ , Launay et al., 2018).

To assess divergent validity, parents also completed the French version of the ADHD-RS (Mercier et al., 2016), a parental questionnaire which assesses inattention and impulsivity symptoms. The ADHD-RS contains 18 child behaviors, whose frequency has to be evaluated by parents. Its French version showed a good factorial structure and dimensions internal consistency (Mercier et al., 2016). Each question is scored 0 to 3 and the total score varies between 0 and 54.

### **Procedure**

Children were tested individually in a quiet office of the hospital. Efforts were made to let children progress through the task at their own pace (as in Experiment 1, we did not over-monitor what they were doing). Since this study was part of a larger research project, the tablet-PST was proposed together with other cognitive tasks assessing pragmatic skills. Tasks were presented in an individualized order with as many pauses as needed in order to adapt to each child's level of attention and engagement, but the sequence of items within the tablet-PST did not vary from Experiment 1.

### **Analysis**

Our analyses involved three steps. First, we assessed convergent and divergent validity of the tablet-PST. Convergent validity was assessed by calculating the correlation between the mean score on TOM items and the EVALEO false-belief raw score. Divergent validity was assessed by calculating the correlation between the mean score on TOM items and the ADHR-RS questionnaire. Spearman rank correlations were used, given the sample size.

Second, we analyzed raw by-trial data at the group level with mixed effect models, by adding the autistic group to the control sample of Experiment 1. As the limited sample size of the autistic group inflated the risk of overfitting and convergence issues, we used the minimally needed models to observe the specific group effects that were of interest. We typically dropped SES which was not implicated in interactions with item-type in Experiment 1, and which did not differ across groups. For both models, we specified group, item-type and age (and their interactions) as fixed effects. We also included random intercepts for items and participants. We then used backward elimination and removed the terms that were not significant from this initial model, in order to reduce model complexity. The model was fitted on all data points for accuracy. For the analysis of response times, we analyzed log-transformed response time data after considering, in line with Langdon & Coltheart (1999), successful responses only (score = 6) in order to reduce complexity in the model. We removed outliers that deviated by more than 2.5 SD in either direction from each group's mean (individually personalized thresholds were not possible due to the exclusion of several items, see Results). Models were fit with the *lme4* package (Bates et al., 2015), sum contrasts were used for all factors except group, where the neurotypical group was used as baseline, and continuous predictors were centered on the mean. Effects were assessed with Satterthwaite's degrees of freedom from the *lmerTest* package (Kuznetsova et al., 2017), and post-hoc contrasts were computed with the *emmeans* package (Lenth, 2022). The effects that did not involve group have already been described in detail in Experiment 1; for this reason and for clarity, we only reported the group effects and interactions that were of interest, while the complete results of the

analysis are displayed in supplementary materials.

Third, we computed summary measures for each participant, according to the procedure of Experiment 1, and discrepancy indices based on the normative sample. That is, we used the linear regressions fitted in Experiment 1 to predict each participant's performance and compared it to their actual performance. To qualify this discrepancy, we followed Crawford & Garthwaite's (2006) methodology, which was shown to be more robust than using the basic model standard error, using the R script developed by Arcara (Montemurro et al., 2022). This discrepancy can be expressed as an estimated percentile rank that ranges from 0 to 100. For comparability, the RT index was reversed, so that a lower estimated percentile rank is always indicative of greater effort or difficulty. We then compared the two groups' standardized scores for the three indices with t-tests and visual examination of the data distributions.

## Results

### Task validation

As expected, the tablet-PST TOM average score was correlated with the EVALEO false belief score ( $\rho = .49, p = .02$ ), indicating convergent validity. Also, as expected, the tablet-PST TOM average score was not significantly correlated with the ADHR-RS total score ( $\rho = .16, p = .50$ ), indicating divergent validity.

### Group-level results

The linear mixed model explaining accuracy (complete output in table S2.2 of supplementary materials) revealed no main

effect of group ( $\beta = -0.09, SE = 0.21, p = .68$ ) but a group by item-type interaction ( $\beta = 0.45, SE = 0.23, p = .046$ ), such that the autistic group performed better than neurotypical children with respect to control items ( $\beta = -0.15, SE = 0.22, p = .50$ ) but worse with respect to TOM items ( $\beta = 0.32, SE = 0.26, p = .21$ ).

In terms of response times, the model was fitted on 1914 data points (after exclusion of 40% of non-perfect responses<sup>26</sup>, then of 1.5% of outliers, see complete output in table S2.3 in supplementary materials), revealing a main effect of group ( $\beta = 0.24, SE = 0.06, p < .001$ ), with autistic participants being slower overall, as well as a three-way group\* item-type \*age ( $\beta = 0.15, SE = 0.07, p < .05$ ) interaction. This interaction revealed that the group difference was not constant over age and item-type (Figure S2.1 in supplementary materials). While TOM items took longer in general than control items, this difference appeared constant over age in the TD group ( $\beta = 0.01, SE = 0.02, p > .05$ ) but tended to decrease in the autistic group ( $\beta = -0.10, SE = 0.05, p = .06$ ). We thus computed group-differences on this item-type contrast at minimum, mean and maximum age, which revealed that the slowdown associated to the TOM items tend to be larger in the autistic group than in the TD group at 6 ( $p = .05$ ) but not at mean or at maximum ages ( $ps > .05$ ).

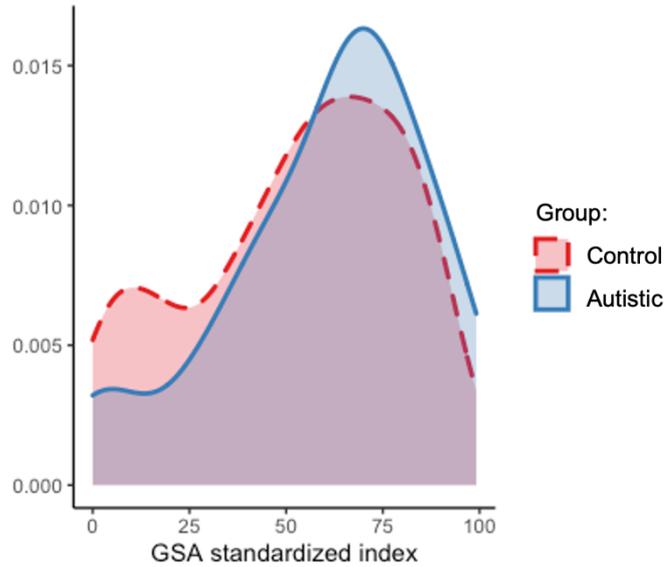
### Individual indices

For each child, we estimated their GSA, TOM-Accuracy and TOM-RT percentile ranks, based on the linear regressions fitted in Experiment 1 (note that the results from the control group in this analysis are retrieved from Experiment 1 so they are not new; they are nevertheless displayed for

---

<sup>26</sup> In line with experiment 1, perfect responses (score = 6) were not distributed equally across items types: they represented 39% of responses

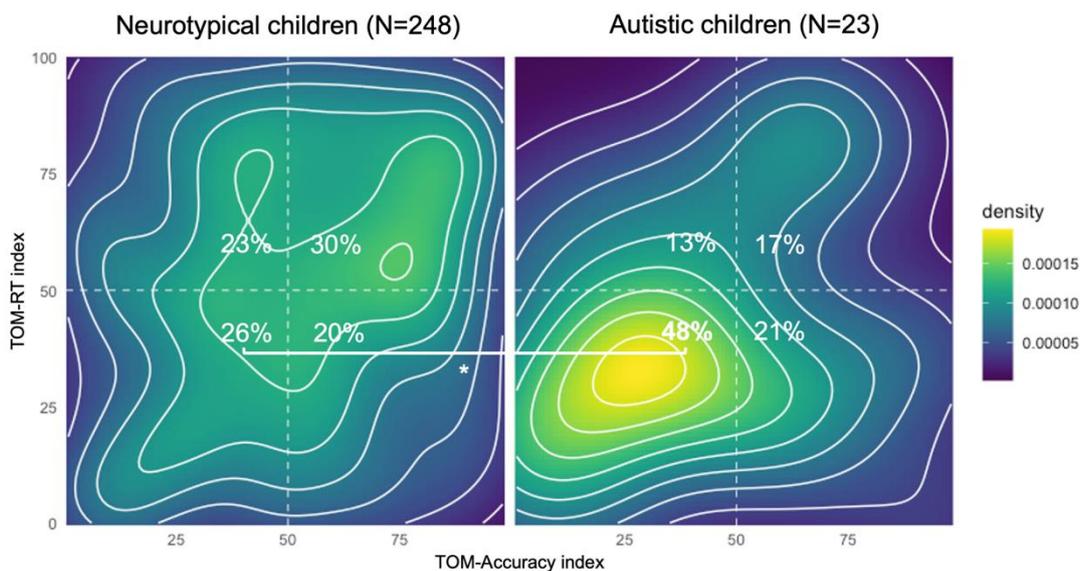
to TOM items and 70% of the responses to non-TOM items



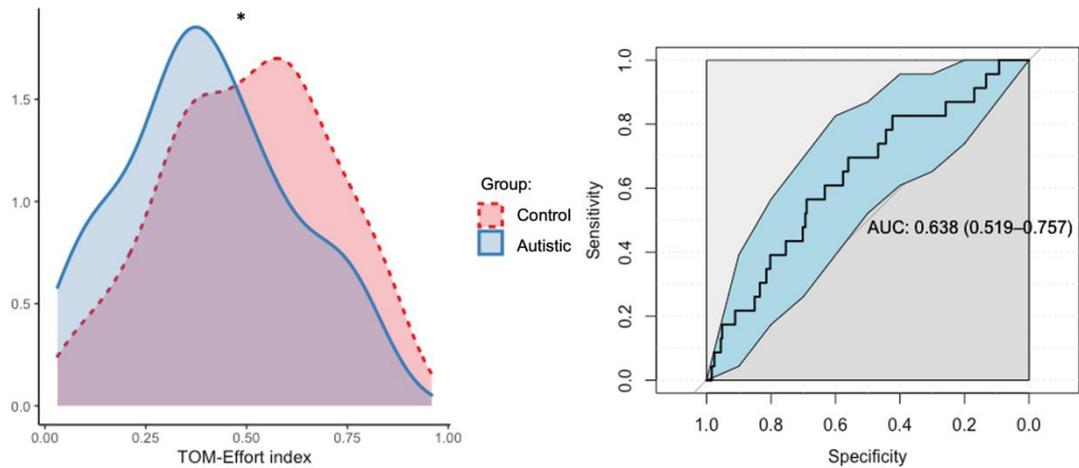
**Figure B-4:** GSA standardized index' distribution for the autistic and the control group

comparison with the autistic group). The Welch two-sample t-tests comparing the groups on those indices failed to reveal significant differences with respect to GSA ( $t(26.5) = -1.32, p = .20$ ), TOM-Accuracy ( $t(25.7) = 1.52, p = .14$ ) and even TOM-RT which approached significance ( $t(27.5) = 1.96, p = .06$ ). However, visual inspection of the distributions suggests interesting effects. On the one hand, it confirms the absence of group-difference for the GSA

(see Figure B-4). On the other hand, it suggests slightly different distributions for the TOM-Accuracy and TOM-RT indices. Specifically, while the control group appears (expectedly) homogeneously distributed, a pattern emerges for the autistic group. As shown in Figure B-5, joint distribution for autistic individuals peaks in the bottom left quarter, where accuracy is poorer, and response-times are longer than expected. Although only one



**Figure B-5:** Joint distributions of the TOM-Accuracy and the TOM-RT indices for the control group (left panel) and the autistic group (right panel).



**Figure B-6:** Distribution across groups (left panel) and ROC curve (right panel) of the TOM-Effort index

fourth of the control group is present in this quarter, half of the autistic sample can be found there ( $X^2(1, N = 271) = 5.10, p < .05$ ).

Based on these results (and following Langdon & Coltheart's, [1999] rationale) we computed a TOM-Effort composite, which is an average of the TOM-Accuracy and TOM-RT indices. This index revealed a significant between-group difference ( $t(26.3) = 2.23, p < .05$ ), which is also clear from visual inspection of the distributions (Figure B-6, left). To examine whether this composite could discriminate between autistic and neurotypical children, we then performed a ROC analysis with this variable (Figure B-6, right). This revealed an area under curve (AUC) of .64 (95% *IC* = [.52-.76]), which was significantly different from chance but indicated only a moderate accuracy.

In [appendix](#), we report on three clinical cases that were extracted from our sample and which illustrate how this task can be used to provide rich clinical information at an individual level.

## Discussion

In Experiment 2, the tablet-PST and the clinical indices built from Experiment 1 were applied to a group of autistic children.

Before discussing the validity of the tablet-PST in the General Discussion, we first briefly discuss the findings from this experiment at the group level in the context of the literature on autism and TOM. First, we note that the gender distribution is quite unbalanced in our autistic sample, which is common in autism research (e.g., Fombonne, 2009). However, Experiment 1 revealed that gender did not affect performance in this task, especially the TOM/non-TOM distinction. Therefore, such unbalance should not obscure the interpretation of the data. An interesting aspect of our results was an interaction between item-type and group on accuracy, such that autistic children performed at higher rates than those in the typically-developing group on control items and at lower rates on false beliefs items. This dissociation is in line with what Baron-Cohen et al. (1986) originally observed with another picture sequencing task in autistic children. It is also consistent with Binnie & Williams (2003), who showed that when asked to sequence pictures, autistic children tended to prefer physical rather than psychological causalities, compared to neurotypical children. This dissociation led to the development of a more general empathizing/systemizing theory of autism (Baron-Cohen, 2009), as a framework to explain the well-described TOM difficulties

of autistic individuals. As for response times, we replicated Kaland et al. (2007), who reported slowdowns on TOM items compared to controls, indicating a more pronounced effort in autism, yet only among younger participants in our case. While our sample size calls for caution regarding the interpretation of such subtle effects and for replication in future studies, it is arguably the case that the tablet-PST is not complex or sensitive enough to reveal such differences among older children. The pattern we observe also echoes the results from Zalla et al. (2006) who observed (with a different picture sequencing task) that autistic children were typically slower in sequencing events involving purposive actions taken by characters than physical events, while it was less the case among the neurotypical participants.

A novel aspect of our study concerns the use of clinical indices to evaluate TOM impairments at the individual level. Although the combined TOM-Effort index was the only one to show a significant difference between the autistic and the control groups, a visual inspection of the distribution of responses reveals a clear pattern in the autistic group that distinguishes itself from the pattern among neurotypical children. That is, TOM items were associated with lower accuracy and slower response times in the autistic group than in neurotypical children. It is interesting that the combined TOM-Effort index revealed itself to be the most discriminating, since combining accuracy and response times has also been identified as a promising method to measure TOM in autistic adults (Livingston et al., 2019). The ROC curve for that measure revealed that it could differentiate the autistic sample from the control group, although with a moderate accuracy (AUC of 64%). This moderate level of accuracy was somewhat expected, and confirms that TOM in general should not be considered *absent* in autism, but

rather partial, delayed or simply atypical, the extent of that difference still being under debate (see Livingston et al., 2019; Marocchini, 2023). Such subtle between-group difference arguably explains that it reaches significance in the combined index only, and not in the accuracy or the RT indices in isolation. Hence, discriminating autistic from non-autistic individuals should not be confounded as a validity measure of whether a given task assesses TOM (Chevallier, 2012).

Finally, Experiment 2 also provided evidence for adequate external validity by showing that the tablet-PST provided results that were consistent (in this group of children *without* language disorder) with a standard false belief task, in line with previous results with the PST in neurotypical children (Rajkumar et al., 2008); moreover, this association seemed to be specifically explained by TOM and not, say, by the general functioning of children, since no correlation appeared with a control measure of ADHD symptoms, as expected.

## General Discussion

This study aimed to assess the suitability of a tablet adaptation of Langdon & Coltheart's (1999)PST in order to ultimately assess children's TOM clinically and individually. This was achieved by testing a large sample of neurotypical children (Experiment 1), providing validation data and allowing for the construction of normalized clinical indices. As a test case, these indices were then applied to a sample of autistic children (Experiment 2).

Overall, our results provided evidence for the validity of the tablet-PST and in several different ways. Structurally, the CFA showed that the task behaved as it was intended. With respect to face and content validity, both the observations on the neurotypical and the autistic groups

confirmed the validity of the task, since it behaved consistently to what can be expected from the TOM and the autism literature and, more specifically, with respect to the literature with paper-and-pencil-based presentations of the PST. Both experiments also supported the tablet-PST's discriminatory power, in the general population (with no floor or ceiling effects in the measures) and in a clinical context. We also showed, with respect to our autistic sample, that the tablet-PST provided consistent results with a standard false-belief task (external validity).

However, standard false-belief tasks typically involve verbal material, which is problematic because language difficulties are frequent in developmental populations. In contrast, the tablet-PST involves minimally verbal material and, as such, is more likely to be accessible to children with language difficulties. Note, however, that we cannot rule out the possibility that participant might employ verbal strategies (e.g., inner speech) to complete the task. Therefore, verbal skills might still play a role in the performance with this task (thought the role of language might be reduced compared to other TOM tasks). Note, however, that this should affect both TOM and non-TOM items, and that using the contrast between these conditions should control for that role, at least partly. Moreover, even when the interference of language in the *performance* in TOM tasks is ruled out, there is evidence supporting an interrelation between language and TOM *competences* (see, e.g., Milligan et al., 2007). In other words, the tablet-PST should probably not be expected to make TOM *independent* from language, but its minimally verbal material makes it suitable to investigate this complex relation in future research (Bosco et al., 2018). In addition to language, this task also controls for another possible confound observed in most TOM tasks, which is working memory (e.g.,

Arslan et al., 2017); indeed, with the tablet-PST all the necessary visual information remains displayed while the item is completed. In terms of reliability, compared to the original PST, the inter-judge reliability of this tablet implementation is expected to be almost perfect, since the entire procedure and its scoring are automatized.

As for applicability, the tablet format of the task may help with data collection. Neurotypical children frequently reported that carrying out the task was enjoyable, and their teachers described the children as especially focused. This was also the case for autistic children, who remained engaged enough to complete the task, even when parents or clinicians had anticipated difficulties. These qualitative observations further support the clinical suitability of the task, and confirmed its usability in group settings (Bignardi et al., 2021).

The literature, as well as our results, shows that children should be expected to perform better on the control items compared to TOM items. However, our indices would allow clinicians to quantify *how much* better and to distinguish different profiles of participants, who might fail TOM items for different reasons. For example, children might fail TOM items because of a general difficulty in picture sequencing, which would be revealed by a low (control) GSA index. But poor performance on TOM items could also result from normal sequencing abilities (attested by a good GSA index) combined with a specific TOM impairment (revealed by a poor TOM-Accuracy index). Similarly, by exploiting response times, our indices can distinguish between a child who performs in what is considered “typical” fashion for this task from another who performs correctly (in terms of accuracy) but exceptionally and effortfully, based on the TOM-RT index. Such reasoning (comparing control and test items, on the one hand, as well as accuracy and response

times, on the other) is generally held by clinicians, and is here directly computed by algorithms. As a consequence, the result of that reasoning presents a reduced risk of human-based bias or between-clinician variability. This is obviously not to say that such calculation replaces clinical reasoning, but that it provides data that is ultimately finer and more patient-calibrated than what raw performance scores could provide alone.

In terms of clinical utility, such data should help drawing a more accurate cognitive and behavioral profile to inform interventions and assess their effects. When used during diagnosis procedure however, this task should be used with much caution since, as we discussed in experiment 2, the sensibility analysis showed that autistic children should not be expected to systematically fail the tablet-PST, nor should neurotypical children be expected to systematically succeed. For this reason, no specific clinical threshold was proposed.

We would be remiss if we did not underline three limitations as we consider future research. The first is that, while we had a large sample in Experiment 1, the sample size of our clinical group was limited. This is why Experiment 2 calls for replication and extension, including to other clinical groups, such as children with language disorders who usually show reduced TOM performance compared to neurotypical children, as shown in Nilsson & De López' (2016) meta-analysis. However, as the authors of this study discuss, the extent to which the verbal nature of the tasks used contributes to this difference is still an open question. Future research should investigate this issue and might benefit from a task like the tablet-PST. Second, in Experiment 2 we compared the clinical sample to the neurotypical children who provided the normative data. Further validation would be provided by comparing this group to another, independent group of neurotypical

children. Third, the cross-sectional methodology we adopted did not allow for test-retest reliability assessment, which could be provided by a multiple testing or longitudinal paradigm.

In sum, after years of research on theory of mind using Langdon & Coltheart's (1999) PST, our study investigates how a tablet-based implementation of this task could be naturally applied to individual clinical assessments. Our data, both in a large neurotypical sample and in a group of autistic children, supports the structural, face and content validity of the tablet-PST, as well as its external validity, sensitivity and clinical relevance. The results illustrate how this task can be helpful to assess the profile of children with atypical development, while providing normative data as background for it. This should benefit the assessment of autistic children but also other conditions that might encounter TOM difficulties, such as those with language disorders or psychiatric conditions.

## References

- Arslan, B., Hohenberger, A., & Verbrugge, R. (2017). Syntactic Recursion Facilitates and Working Memory Predicts Recursive Theory of Mind. *PLOS ONE*, *12*(1), e0169510. <https://doi.org/10.1371/journal.pone.0169510>
- Baltazar, M., Geoffray, M., Chatham, C., Bouvard, M., Martinez Teruel, A., Monnet, D., Scheid, I., Murzi, E., Couffin-Cadiergues, S., Umbricht, D., Murtagh, L., Delorme, R., Ly Le-Moal, M., Leboyer, M., & Amestoy, A. (2021). "Reading the Mind in the Eyes" in Autistic Adults is Modulated by Valence and Difficulty: An InFoR Study. *Autism Research*, *14*(2), 380–388. <https://doi.org/10.1002/aur.2390>
- Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *TRENDS in Cognitive Sciences*, 6(6), 248–254.
- Baron-Cohen, S. (2009). Autism: The Empathizing–Systemizing (E-S) Theory. *Annals of the New York Academy of Sciences*, 1156(1), 68–80. <https://doi.org/10.1111/j.1749-6632.2009.04467.x>
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and Intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology*, 4(2), 113–125. <https://doi.org/10.1111/j.2044-835X.1986.tb01003.x>
- Baron-Cohen, S., O’Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). *Recognition of Faux Pas by Normally Developing Children and Children with Asperger Syndrome or High-Functioning Autism*. 12.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *J. Child Psychol. Psychiat.*, 42(2), 241–251.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/doi:10.18637/jss.v067.i01>.
- Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic Review and Inventory of Theory of Mind Measures for Young Children. *Frontiers in Psychology*, 10, 2905. <https://doi.org/10.3389/fpsyg.2019.02905>
- Behrmann, M., Avidan, G., Leonard, G. L., Kimchi, R., Luna, B., Humphreys, K., & Minshew, N. (2006). Configural processing in autism and its relationship to face processing. *Neuropsychologia*, 44(1), 110–129. <https://doi.org/10.1016/j.neuropsychologia.2005.04.002>
- Bignardi, G., Dalmaijer, E. S., Anwyll-Irvine, A., & Astle, D. E. (2021). Collecting big data with small screens: Group tests of children’s cognition with touchscreen tablets are reliable and valid. *Behavior Research Methods*, 53(4), 1515–1529. <https://doi.org/10.3758/s13428-020-01503-3>
- Binnie, L., & Williams, J. (2003). Intuitive Psychology and Physics among Children with Autism and Typically Developing Children. *Autism*, 7(2), 173–193. <https://doi.org/10.1177/1362361303007002005>
- Bivona, U., Formisano, R., Laurentiis, S., Accetta, N., Cosimo, M., Massicci, R., Ciurli, P., Azicnuda, E., Silvestro, D., Sabatini, U., Falletta Caravasso, C., Carlesimo, G., Caltagirone, C., & Costa, A. (2015). Theory of mind impairment after severe traumatic brain injury and its relationship with caregivers’ quality of life. *Restorative Neurology and Neuroscience*, 33. <https://doi.org/10.3233/RNN-140484>
- Blair, C., & Razza, R. P. (2007). Relating Effortful Control, Executive Function, and False Belief Understanding to Emerging Math and Literacy Ability in Kindergarten. *Child Development*, 78(2), 647–663. <https://doi.org/10.1111/j.1467-8624.2007.01019.x>
- Bosco, F. M., Gabbatore, I., & Tirassa, M. (2014). A broad assessment of theory of mind in adolescence: The complexity of mindreading. *Consciousness and Cognition*, 24, 84–97. <https://doi.org/10.1016/j.concog.2014.01.003>
- 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>
- Botha, M., Hanlon, J., & Williams, G. L. (2021). Does Language Matter? Identity-First Versus Person-First Language Use in Autism Research: A Response to Vivanti. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-020-04858-w>
- Botting, N., & Conti-Ramsden, G. (2008). The role of language, social cognition, and social skill in the functional social outcomes of young adolescents with and without a history of SLI. *British Journal of Developmental Psychology*, 26(2), 281–300. <https://doi.org/10.1348/026151007X235891>
- Bury, S. M., Jellett, R., Spoor, J. R., & Hedley, D. (2020). “It defines who I am” or “It’s

- something I have”: What language do [autistic] Australian adults [on the autism spectrum] prefer? *Journal of Autism and Developmental Disorders*, 1–11.
- Charman, T., Ruffman, T., & Clements, W. (2002). Is there a Gender Difference in False Belief Development? *Social Development*, 11(1), 1–10. <https://doi.org/10.1111/1467-9507.00183>
- Chevallier, C. (2012). Theory of mind and autism: Revisiting Baron-Cohen et al.’s Sally-Anne study. In *Developmental psychology: Revisiting the classic studies* (pp. 148–163). Sage Publications Ltd.
- Crawford, J. R., & Garthwaite, P. H. (2006). Comparing patients’ predicted test scores from a regression equation with their obtained scores: A significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology*, 20(3), 259–271. <https://doi.org/10.1037/0894-4105.20.3.259>
- Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) Family Affluence Scale. *Social Science & Medicine*, 66(6), 1429–1436. <https://doi.org/10.1016/j.socscimed.2007.11.024>
- de Villiers, J. (2007). The interface of language and Theory of Mind. *Lingua*, 117(11), 1858–1878. <https://doi.org/10.1016/j.lingua.2006.11.006>
- Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1(4), 568–570. <https://doi.org/10.1017/S0140525X00076664>
- Dunn, J., Cutting, A. L., & Fisher, N. (2002). Old Friends, New Friends: Predictors of Children’s Perspective on Their Friends at School. *Child Development*, 73(2), 621–635. <https://doi.org/10.1111/1467-8624.00427>
- Fink, E., Begeer, S., Hunt, C., & de Rosnay, M. (2014). False-Belief Understanding and Social Preference Over the First 2 Years of School: A Longitudinal Study. *Child Development*, n/a-n/a. <https://doi.org/10.1111/cdev.12302>
- Fletcher-Watson, S., & Happé, F. (2019). *Autism: A new introduction to psychological theory and current debate*. Routledge.
- Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research*, 65(6), 591–598.
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431), 459–473. <https://doi.org/10.1098/rstb.2002.1218>
- Happé, F. (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. *Journal of Autism and Developmental Disorders*, 24(2), 129–154. <https://doi.org/10.1007/BF02172093>
- Hayward, E. O., & Homer, B. D. (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology*, 35(3), 454–462.
- Jurek, L., Longuet, Y., Baltazar, M., Amestoy, A., Schmitt, V., Desmurget, M., & Geoffroy, M.-M. (2019). How did I get so late so soon? A review of time processing and management in autism. *Behavioural Brain Research*, 374, 112121. <https://doi.org/10.1016/j.bbr.2019.112121>
- Kaland, N., Smith, L., & Mortensen, E. L. (2007). Response Times of Children and Adolescents with Asperger Syndrome on an ‘Advanced’ Test of Theory of Mind. *Journal of Autism and Developmental Disorders*, 37(2), 197–209. <https://doi.org/10.1007/s10803-006-0152-8>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26.
- Lane, R., & Radesky, J. (2019). Digital Media and Autism Spectrum Disorders: Review of Evidence, Theoretical Concerns, and Opportunities for Intervention. *Journal of Developmental and Behavioral Pediatrics: JDBP*, 40(5), 364–368. <https://doi.org/10.1097/DBP.0000000000000664>

- Langdon, R., & Coltheart, M. (1999). Mentalising, schizotypy, and schizophrenia. *Cognition*, 71(1), 43–71. [https://doi.org/10.1016/S0010-0277\(99\)00018-9](https://doi.org/10.1016/S0010-0277(99)00018-9)
- Launay, L., Maeder, C., Roustit, J., & Touzin, M. (2018). *Évaluation du Langage écrit et du Langage Oral 6-15 ans (EVALEO 6-15)*. OrthoÉdition.
- Lenth, R. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (R package version 1.8.2). [Computer software]. <https://CRAN.R-project.org/package=emmeans>
- Levy, S. E., Giarelli, E., Lee, L.-C., Schieve, L. A., Kirby, R. S., Cunniff, C., Nicholas, J., Reaven, J., & Rice, C. E. (2010). Autism Spectrum Disorder and Co-occurring Developmental, Psychiatric, and Medical Conditions Among Children in Multiple Populations of the United States. *Journal of Developmental & Behavioral Pediatrics*, 31(4), 267. <https://doi.org/10.1097/DBP.0b013e3181d5d03b>
- Livingston, L. A., Carr, B., & Shah, P. (2019). Recent Advances and New Directions in Measuring Theory of Mind in Autistic Adults. *Journal of Autism and Developmental Disorders*, 49(4), 1738–1744. <https://doi.org/10.1007/s10803-018-3823-3>
- Livingston, L. A., & Happé, F. (2017). Conceptualising compensation in neurodevelopmental disorders: Reflections from autism spectrum disorder. *Neuroscience & Biobehavioral Reviews*, 80, 729–742. <https://doi.org/10.1016/j.neubiorev.2017.06.005>
- Maat, A., Fett, A.-K., & Derks, E. (2012). Social cognition and quality of life in schizophrenia. *Schizophrenia Research*, 137(1), 212–218. <https://doi.org/10.1016/j.schres.2012.02.017>
- Marocchini, E. (2023). Impairment or difference? The case of Theory of Mind abilities and pragmatic competence in the Autism Spectrum. *Applied Psycholinguistics*, 1–19. <https://doi.org/10.1017/S0142716423000024>
- Mercier, C., Roche, S., Gaillard, S., Kassai, B., Arzimanoglou, A., Herbillon, V., Roy, P., & Rheims, S. (2016). Partial validation of a French version of the ADHD-rating scale IV on a French population of children with ADHD and epilepsy. Factorial structure, reliability, and responsiveness. *Epilepsy & Behavior*, 58, 1–6. <https://doi.org/10.1016/j.yebeh.2016.02.016>
- Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind: Meta-Analysis of the Relation Between Language Ability and False-belief Understanding. *Child Development*, 78(2), 622–646. <https://doi.org/10.1111/j.1467-8624.2007.01018.x>
- Montemurro, S., Daini, R., Tagliabue, C., Guzzetti, S., Gualco, G., Mondini, S., & Arcara, G. (2022). Cognitive reserve estimated with a life experience questionnaire outperforms education in predicting performance on MoCA: Italian normative data. *Current Psychology*. <https://doi.org/10.1007/s12144-022-03062-6>
- Nilsson, K. K., & De López, K. J. (2016). Theory of Mind in Children With Specific Language Impairment: A Systematic Review and Meta-Analysis. *Child Development*, 87(1), 143–153. <https://doi.org/10.1111/cdev.12462>
- O’Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A Clinical Assessment Tool for Advanced Theory of Mind Performance in 5 to 12 Year Olds. *Journal of Autism and Developmental Disorders*, 39(6), 916–928. <https://doi.org/10.1007/s10803-009-0699-2>
- Payne, J. M., Porter, M., Pride, N. A., & North, K. N. (2016). Theory of mind in children with Neurofibromatosis Type 1. *Neuropsychology*, 30(4), 439–448. <https://doi.org/10.1037/neu0000262>
- Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that...” attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, 39(3), 437–471. [https://doi.org/10.1016/0022-0965\(85\)90051-7](https://doi.org/10.1016/0022-0965(85)90051-7)
- Piatt, C., Coret, M., Choi, M., Volden, J., & Bisanz, J. (2016). Comparing Children’s Performance on and Preference for a Number-Line Estimation Task: Tablet Versus Paper and Pencil. *Journal of Psychoeducational Assessment*, 34(3), 244–255. <https://doi.org/10.1177/0734282915594746>

- Porter, M. A., Coltheart, M., & Langdon, R. (2008). Theory of Mind in Williams Syndrome Assessed Using a Nonverbal Task. *Journal of Autism and Developmental Disorders*, 38(5), 806–814. <https://doi.org/10.1007/s10803-007-0447-4>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515–526. <https://doi.org/10.1017/S0140525X00076512>
- Pyers, J. E., & Senghas, A. (2009). Language Promotes False-Belief Understanding: Evidence From Learners of a New Sign Language. *Psychological Science*, 20(7), 805–812. <https://doi.org/10.1111/j.1467-9280.2009.02377.x>
- R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rajkumar, A. P., Yovan, S., Raveendran, A. L., & Russell, P. S. S. (2008). Can only intelligent children do mind reading: The relationship between intelligence and theory of mind in 8 to 11 years old. *Behavioral and Brain Functions*, 4(1), 51. <https://doi.org/10.1186/1744-9081-4-51>
- Rosseel, Y. (2012). **lavaan**: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>
- Senju, A. (2012). Spontaneous Theory of Mind and Its Absence in Autism Spectrum Disorders. *The Neuroscientist*, 18(2), 108–113. <https://doi.org/10.1177/1073858410397208>
- Van Rheenen, T. E., & Rossell, S. L. (2013). Picture sequencing task performance indicates theory of mind deficit in bipolar disorder. *Journal of Affective Disorders*, 151(3), 1132–1134. <https://doi.org/10.1016/j.jad.2013.07.009>
- Van 'T Hof, M., Tisseur, C., Van Berckeleer-Onnes, I., Van Nieuwenhuyzen, A., Daniels, A. M., Deen, M., Hoek, H. W., & Ester, W. A. (2021). Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. *Autism*, 25(4), 862–873. <https://doi.org/10.1177/1362361320971107>
- Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, 14(8), 383–388. <https://doi.org/10.1016/j.tics.2010.05.006>
- Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, 15(6), 728–755. <https://doi.org/10.1080/17405629.2018.1435413>
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, 72(3), 655–684. <https://doi.org/10.1111/1467-8624.00304>
- Westby, C., & Robinson, L. (2014). A developmental perspective for promoting theory of mind. *Topics in Language Disorders*, 34(4), 362–382.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.
- Zalla, T., Labruyere, N., & Georgieff, N. (2006). Goal-Directed Action Representation in Autism. *Journal of Autism and Developmental Disorders*, 36(4), 527–540. <https://doi.org/10.1007/s10803-006-0092-3>

## Appendix: clinical cases

In order to illustrate how the tablet-PST can be used to provide clinical information, richer than a simple standardized score on a classical task, we extracted three prototypical profiles from our sample.



Participant 1 is a 6;7 years old boy. His autism diagnostic was confirmed by a positive SCQ and a positive ADOS-2. His level of intelligence functioning was also assessed and is strictly normal. In the tablet-PST, he obtained a GSA of 2.8/6, placing him only at the 1<sup>st</sup> percentile, given his age and SES. This score indicates that the task in itself is not suitable to assess this specific child's skills. Indeed, the ability of that child to sequence picture is compromised, maybe due to his young age, to behavioral difficulties and/or to difficulties in sequencing events and representing time which can be observed in certain individuals with autism (Jurek et al., 2019). If theory of mind must be assessed, it should be done otherwise.



Participant 7 presents a different profile. She is a 7;6 years old girl, whose autism diagnostic was confirmed by a positive SCQ and a positive ADI-R. Her level of intelligence functioning hasn't been formally tested, but no difficulties were suspected by the clinicians who follow her, and her language skills were in the normal to normal-sup range. In the tablet-PST, she got a 4.9/6 GSA (71<sup>st</sup> percentile) but scored 1.5/6 on average on TOM items (3<sup>rd</sup> percentile). Contrary to participant 1, participant 7 is thus very able to sequence pictures in general, with a GSA that is even in the normal-superior range. However, specific difficulties arise as soon as TOM is involved, where she scores much lower than expected, given her age and her good sequencing abilities. This is indicative of specific difficulties in theory of mind. No particularity was observed in her response times: she answered TOM items in 29 seconds in average, contra 23 seconds for the control items, which places her at percentile 26.



Finally, participant 6 is a 7;6 years old boy, whose diagnostic was confirmed by a positive ADOS-2 and a positive SCQ. His intelligence was assessed as normal, as well as his language skills. He performed very well on the tablet-PST control, items with a GSA of 5.6/6 (85<sup>th</sup> percentile), and even better on TOM items with a mean accuracy of 6/6 (94<sup>th</sup> percentile). This child has no problem in sequencing picture in general and, given his age and his good GSA, he performs even significantly better than expected in TOM items. However, he took 44 seconds in average to answer to items involving false

beliefs, contra only 18 seconds on control items: this difference, given his good accuracy, places him only at the 1<sup>st</sup> percentile. This indicated that participant 6's good accuracy on TOM items is obtained through a clearly unusually important effort, in terms of response times, that is specific to items where TOM was involved. This might reflect atypical processing strategies such as compensatory mechanisms specifically associated with mentalizing.

## Supplementary materials

### EXPERIMENT 1

**Table S1.1:** Cronbach alpha value after removing each item (raw alpha = .75)

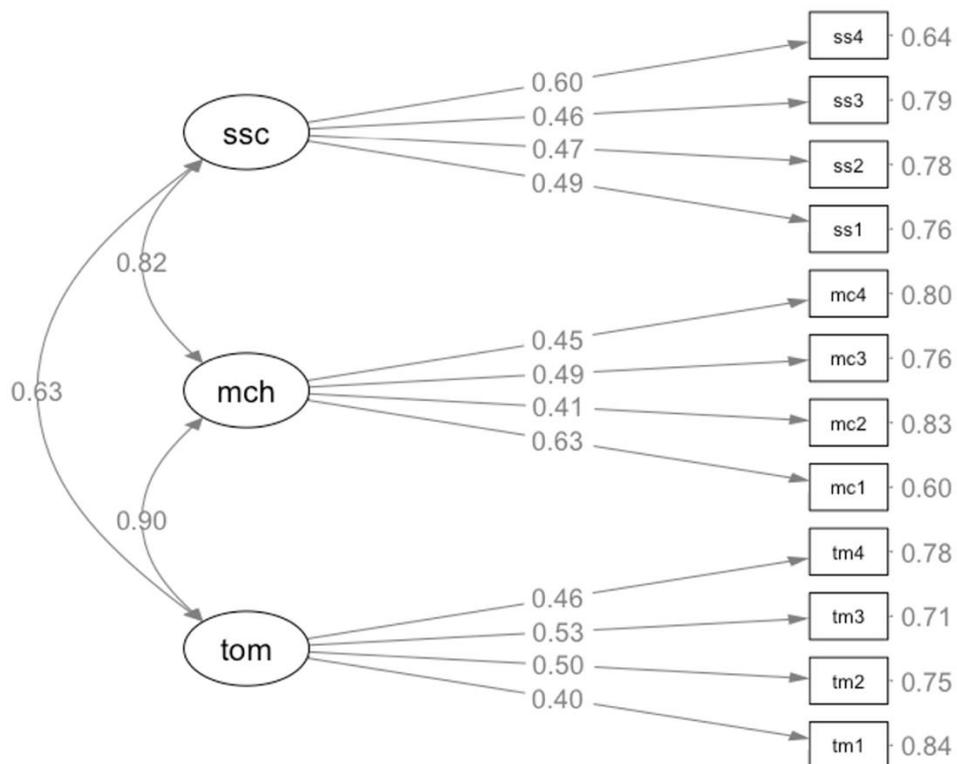
<i>Item</i>	<i>Alpha if deleted</i>	<i>Item</i>	<i>Alpha if deleted</i>	<i>Item</i>	<i>Alpha if deleted</i>
<b>mechanical1</b>	.72	<b>sscript1</b>	.73	<b>tom1</b>	.74
<b>mechanical2</b>	.73	<b>sscript2</b>	.73	<b>tom2</b>	.73
<b>mechanical3</b>	.72	<b>sscript3</b>	.74	<b>tom3</b>	.73
<b>mechanical4</b>	.73	<b>sscript4</b>	.72	<b>tom4</b>	.73

**Table S1.2:** Estimated parameters of the model explaining accuracy while considering separately mechanical and social-scripts items

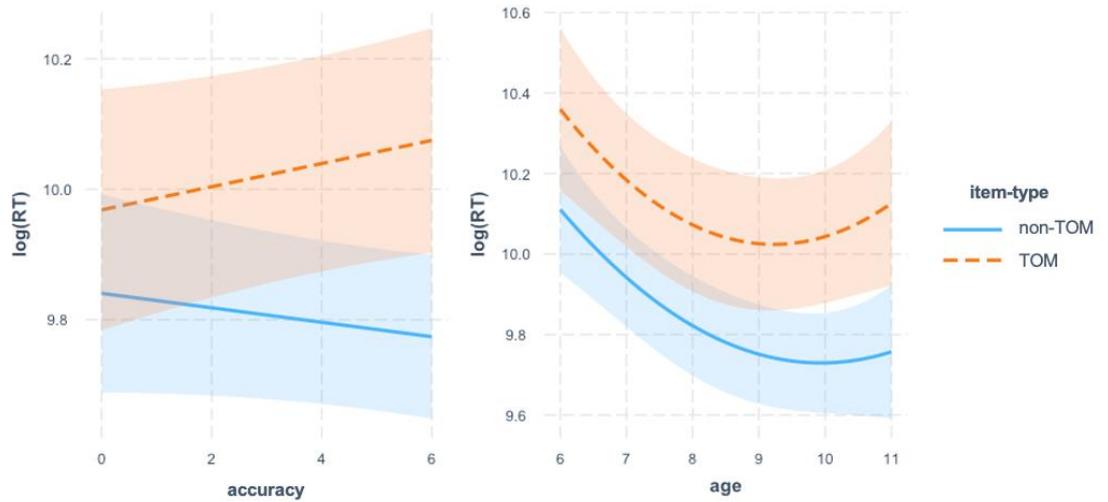
<i>Parameter</i>	<i>Coefficient</i>	<i>Standard error</i>	<i>p</i>
<b>(Intercept)</b>	3.30	0.26	< .001
<b>item-type1</b>	0.14	0.80	.866
<b>item-type2</b>	1.76	0.80	.028
<b>gender</b>	0.08	0.11	.489
<b>age</b>	21.24	3.03	< .001
<b>age<sup>2</sup></b>	-4.39	3.02	.146
<b>SES</b>	0.02	0.00	< .001
<b>item-type1 * gender</b>	-0.23	0.27	.405
<b>item-type2 * gender</b>	0.55	0.27	.044
<b>item-type1 * age</b>	12.49	7.40	.092
<b>item-type2 * age</b>	-22.40	7.40	.003
<b>item-type1 * age<sup>2</sup></b>	9.25	7.39	.211
<b>item-type2 * age<sup>2</sup></b>	7.93	7.39	.283
<b>item-type1 * SES</b>	0.01	0.01	.180
<b>item-type2 * SES</b>	-0.00	0.01	.564

**Chapitre 8 • Table S1.3:** Estimated parameters of the model explaining log-transformed response times while considering separately mechanical and social-scripts items

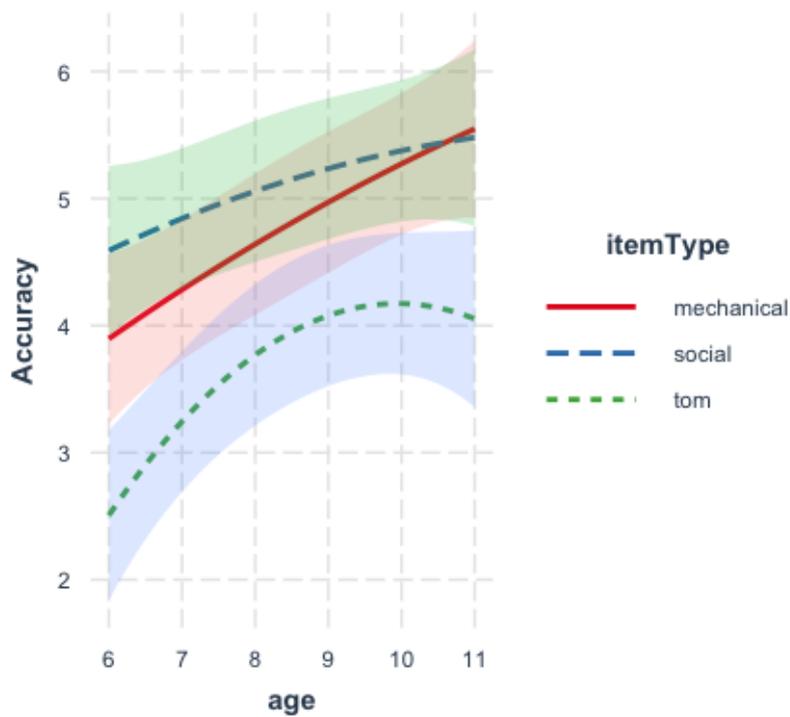
<i>Parameter</i>	<i>Coefficient</i>	<i>Standard error</i>	<i>p</i>
<i>(Intercept)</i>	9.77	0.08	< .001
<b>item-type1</b>	0.09	0.23	.680
<b>item-type2</b>	-0.22	0.23	.345
<b>score</b>	0.01	0.00	.107
<b>SES</b>	0.00	0.00	.175
<b>age</b>	-4.83	0.98	< .001
<b>age<sup>2</sup></b>	2.45	0.98	.012
<b>gender</b>	0.07	0.04	.058
<b>item-type1 * score</b>	-0.03	0.02	.080
<b>item-type2 * score</b>	-0.04	0.02	.027
<b>item-type1 * SES</b>	-0.00	0.00	.225
<b>item-type2 * SES</b>	-0.00	0.00	.886
<b>item-type1 * age</b>	-2.72	1.59	.088
<b>item-type2 * age</b>	-3.27	1.58	.038
<b>item-type1 * age<sup>2</sup></b>	-1.64	1.54	.288
<b>item-type2 * age<sup>2</sup></b>	0.45	1.55	.770
<b>item-type1 * gender</b>	0.07	0.06	.226
<b>item-type2 * gender</b>	-0.09	0.06	.137



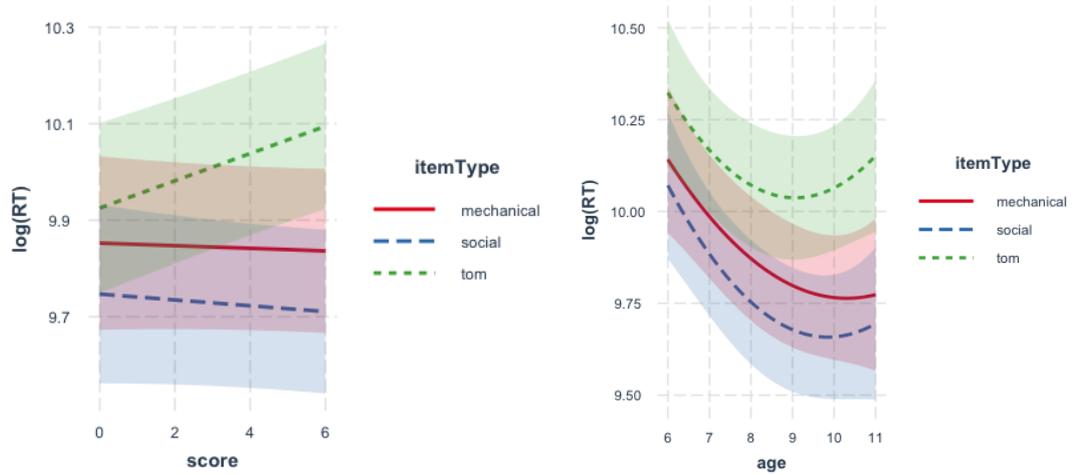
**Figure S1.1:** Confirmatory Factor Analysis of the tablet-PST; each arrow represents a significant relationship (all  $p$ s < .001); values indicate standardized coefficients. SSC = social scripts, MCH = mechanical, TOM = theory of mind.



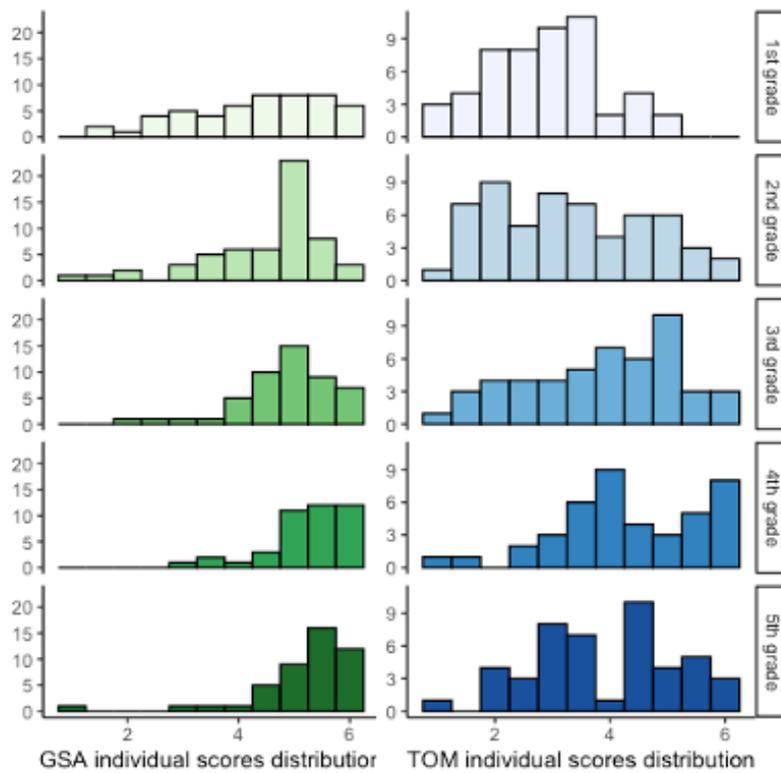
**Figure S1.2:** model's predicted log-transformed response times as a function of item-type and accuracy (left) and age (right); shaded areas represent 95% confidence intervals



**Figure S1.3:** model's predicted accuracy as a function of age and *detailed* item-type, when considering separately mechanical and social-scripts items; shaded areas represent 95% confidence intervals



**Figure S1.4:** model's predicted log-transformed response times as a function of *detailed* item-type and accuracy (left) and age (right), when considering separately mechanical and social-scripts items; shaded areas represent 95% confidence intervals



**Figure S1.5:** Distribution of the two raw composites in the 5 grades-groups

## EXPERIMENT 2

**Table S2.1:** Clinical group detailed description

<b>ID</b>	<b>Age (years; months)</b>	<b>Class</b>	<b>Sex</b>	<b>SES</b>	<b>Autism Assessment</b>	<b>Intelligence assessment*</b>
<b>1</b>	6;7	1 <sup>st</sup> grade	M	80	positive ADOS-2; positive SCQ	FSIQ = 105 (WPPSI-IV)
<b>2</b>	6;7	1 <sup>st</sup> grade	M	17	positive ADOS-2; negative SCQ	<i>NA</i>
<b>3</b>	7;1	1 <sup>st</sup> grade	M	50	positive SCQ; positive ADI-R	VC = 100; VS = 102; FR = 109 (WISC-V)
<b>4</b>	7;2	1 <sup>st</sup> grade	M	20	positive SCQ	FSIQ = 131 (WPPSI-IV)
<b>5</b>	7;2	1 <sup>st</sup> grade	M	65	positive SCQ	FSIQ = 138 (WISC-V)
<b>6</b>	7;7	2 <sup>nd</sup> grade	M	65	positive ADOS-2; positive SCQ	VC = 98; VS = 102; FR = 112 (WISC-V)
<b>7</b>	7;7	2 <sup>nd</sup> grade	F	43	positive SCQ	<i>NA</i>
<b>8</b>	7;8	3 <sup>rd</sup> grade	M	35	positive ADOS-2; positive SCQ	VC = 114; VS = 117; FR = 135 (WPPSI-VI)
<b>9</b>	8;3	2 <sup>nd</sup> grade adapted	F	24	positive ADOS-2; positive SCQ	VC = 92; VS = 89; FR = 97 (WISC-V)
<b>10</b>	8;6	3 <sup>rd</sup> grade	M	49	positive ADOS-2; positive SCQ	VC = 95; VS = 105; FR = 118 (WISC-V)
<b>11</b>	8;6	3 <sup>rd</sup> grade	M	54	positive ADOS-2	FSIQ = 105 (WISC-V)
<b>12</b>	8;9	3 <sup>rd</sup> grade	M	84	positive ADOS-2; positive SCQ	FSIQ = 79 (WPPSI-IV)
<b>13</b>	8;10	4 <sup>th</sup> grade	M	60	positive ADOS-2; negative SCQ	FSIQ = 124 (WPPSI-IV)
<b>14</b>	9;2	3 <sup>rd</sup> grade adapted	M	54	positive ADOS-2	FSIQ = 88 (WNV)
<b>15</b>	9;4	4 <sup>th</sup> grade	M	58	positive ADOS-2; positive SCQ	VC = 98; VS = 102; FR = 112 (WISC-V)
<b>16</b>	9;10	4 <sup>th</sup> grade	M	50	positive ADOS-2; positive SCQ	FSIQ = 92 (WISC-V)
<b>17</b>	10;0	4 <sup>th</sup> grade	F	62	positive ADOS-2; positive SCQ	FSIQ = 100 (WISC-V)
<b>18</b>	10;1	4 <sup>th</sup> grade	M	61	positive ADOS-2; positive SCQ	VC = 118; VS = 138; FR = 118 (WISC-V)
<b>19</b>	10;3	4 <sup>th</sup> grade adapted	M	72	positive SCQ	<i>NA</i>
<b>20</b>	10;4	5 <sup>th</sup> grade	M	47	positive ADOS-2; negative SCQ	FSIQ = 112 (WISC-V)
<b>21</b>	10;6	5 <sup>th</sup> grade	M	56	positive ADOS-2; negative SCQ	FSIQ = 115 (WPPSI-IV)
<b>22</b>	10;7	5 <sup>th</sup> grade	M	64	positive SCQ	FSIQ = 127 (WISC-V)
<b>23</b>	10;8	5 <sup>th</sup> grade	M	54	positive ADOS-2	FSIQ = 121 (WISC-V)

\* FSIQ is reported when available. If not, other available summary values are reported. The tests used to provide these measures are detailed.

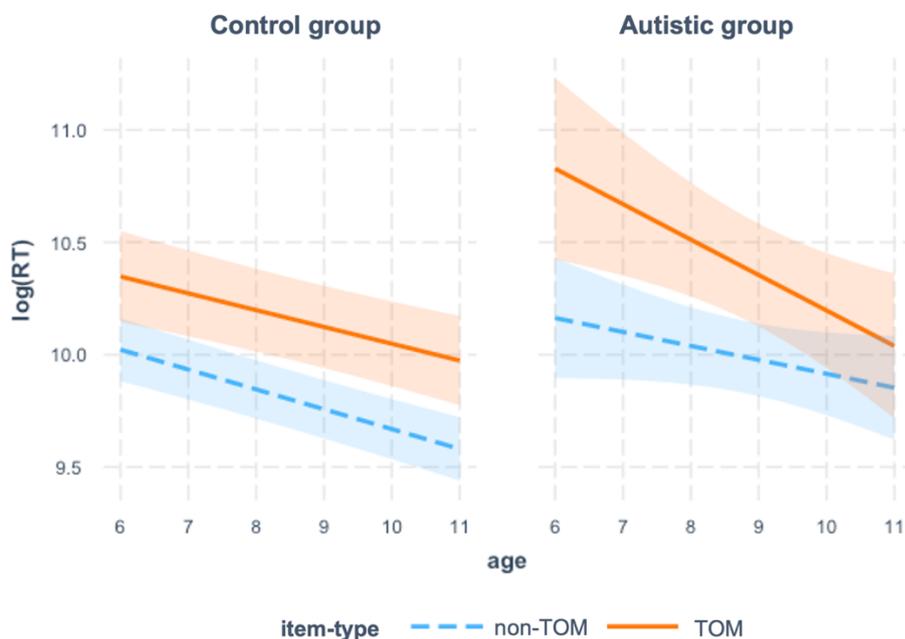
*Abbreviations used:* SCQ = Social Communication Questionnaire; ADOS-2 = Autism Diagnostic Observation Scale 2<sup>nd</sup> Edition; FSIQ = Full Scale Intelligence Quotient; VC = Verbal Comprehension; VS = Visuo-Spatial; FR = Fluid Reasoning; WPPSI-IV = Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition; WISC-V = Wechsler Intelligence Scale for Children, Fifth Edition; WNV= Wechsler Nonverbal Scale of Ability

**Table S2.2:** Estimated parameters of the model explaining accuracy in Experiment 2

<b>Parameter</b>	<b>Coefficient</b>	<b>Standard error</b>	<b>p</b>
(Intercept)	4.28	0.16	< .001
group [autistic]	-0.09	0.21	.678
<b>item-type</b>	<b>1.14</b>	<b>0.30</b>	<b>&lt; .001</b>
<b>age</b>	<b>0.37</b>	<b>0.06</b>	<b>&lt; .001</b>
<b>group [autistic] * item-type</b>	<b>0.46</b>	<b>0.23</b>	<b>.046</b>

**Table S2.3:** Estimated parameters of the model explaining log-transformed response times in Experiment 2

<b>Parameter</b>	<b>Coefficient</b>	<b>Standard error</b>	<b>p</b>
(Intercept)	9.97	0.06	< .001
<b>item-type</b>	<b>-0.36</b>	<b>0.11</b>	<b>.001</b>
<b>group</b>	<b>0.24</b>	<b>0.06</b>	<b>&lt; .001</b>
<b>age</b>	<b>-0.11</b>	<b>0.02</b>	<b>&lt; .001</b>
item-type * group	-0.06	0.07	.446
item-type * age	-0.02	0.02	.386
group * age	-0.04	0.06	.512
<b>(item-type * group) * age</b>	<b>0.15</b>	<b>0.07</b>	<b>.042</b>



**Figure S2.1:** Three-way interaction between age, group, and item-type on the response times of succeeded items (shaded areas represent 95% confidence intervals)

## Annexe C • Matériel supplémentaire pour l'étude 2

### Experiment 1

**Table S1.1:** *Vignette characteristics*

	<i>Metaphor-base</i>	<i>Synonymic-base</i>	<i>Stat. comparison</i>
Mean length in words (SD)	54.1 (4.0)	54.0 (4.1)	Z = -0.2, p > .05
Mean readability (SD) <sup>1</sup>	52.4 (8.6)	53.0 (6.8)	Z = -0.4, p > .05
Mean content words frequency (SD) <sup>2</sup>	5.7 (0.5)	5.3 (0.4)	Z = -1.3, p > .05
Mean referent frequency (SD) <sup>3</sup>	59.1 (7.0)	56.9 (9.2)	Z = -0.5, p > .05
Mean target noun frequency (SD) <sup>3</sup>	51.5 (10.3)		/

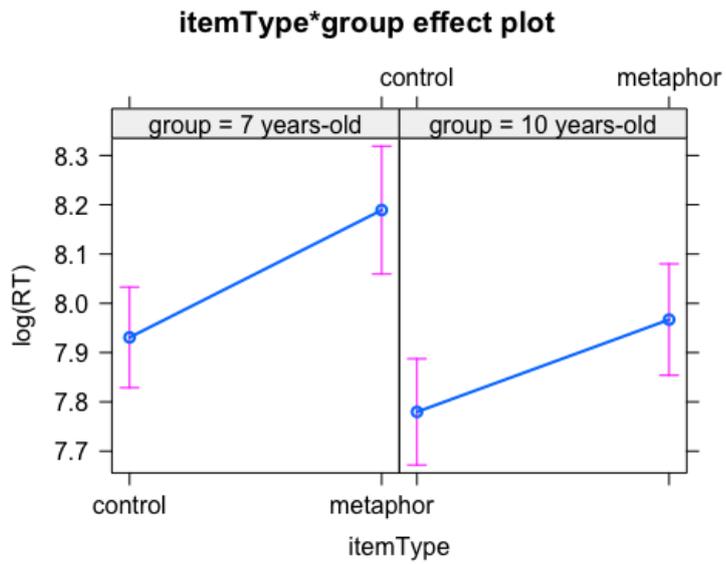
<sup>1</sup> As indexed by the French tool Scolarius ([www.scolarius.com](http://www.scolarius.com)); a high score indicates low readability and scores ranging from 50 to 89 are considered suitable for primary school pupils.

<sup>2</sup> As indexed by the French movie-subtitles database Lexique 2 (New et al., 2004); the median frequency was high, at 297 per million occurrences; this value is at the 97<sup>th</sup> percentile of the entire database. Since frequencies were not distributed normally, we used the log-transformed data to assess difference across conditions.

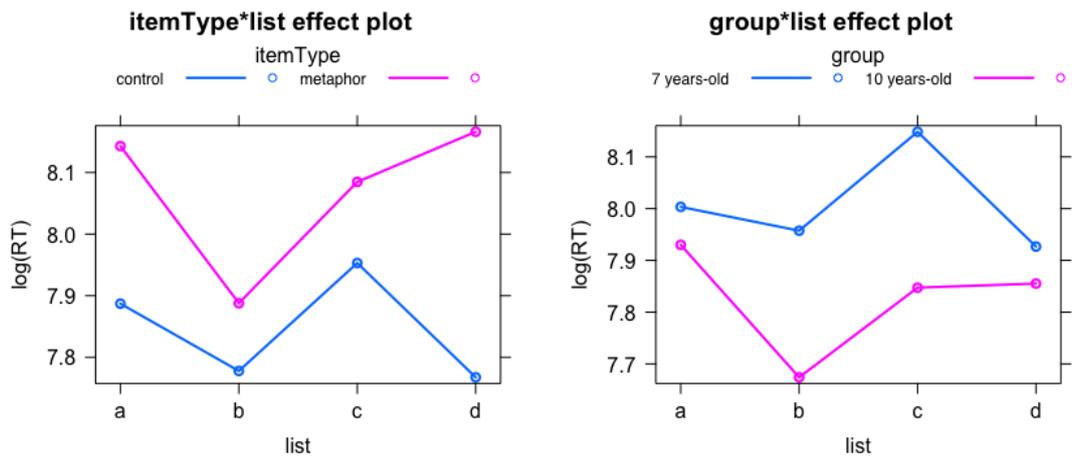
<sup>3</sup> As measured in the Manulex database (Lété, 2003) which gathers French Textbooks for primary schools (1<sup>st</sup> to 5<sup>th</sup> grade). The overall interquartile range in this database is [24-46].

**Table S1.2:** *By-item results*

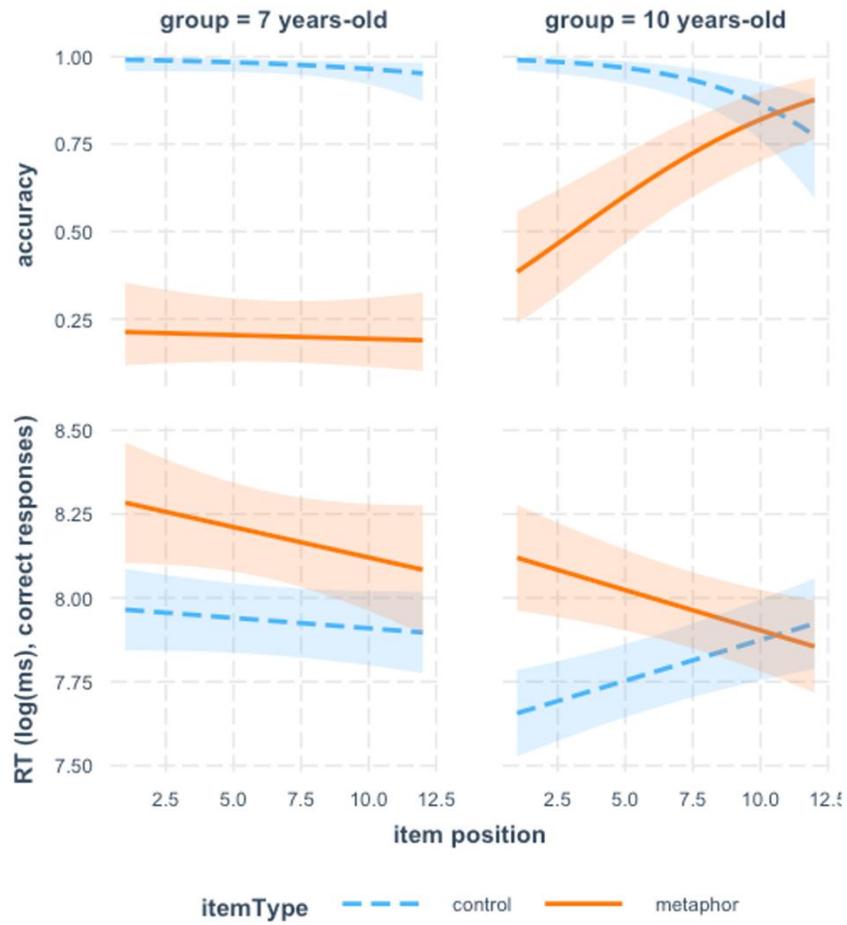
Item	Prop. of metaphoric interpretations in the whole sample		Prop. of metaphoric interpretations in the metaphor condition		Mean log-transformed RT (SD)	
	Literal context	Metaphor context	7 years-old	10 years-old	Literal context	Metaphoric context
<b>1</b>	0%	46%	18%	71%	7,73 (0,48)	8,01 (0,42)
<b>2</b>	9%	36%	18%	55%	7,92 (0,40)	8,09 (0,38)
<b>3</b>	0%	42%	23%	71%	7,85 (0,44)	7,87 (0,42)
<b>4</b>	11%	36%	18%	55%	7,72 (0,50)	8,10 (0,39)
<b>5</b>	2%	42%	27%	65%	8,07 (0,36)	8,36 (0,29)
<b>6</b>	0%	38%	27%	52%	7,58 (0,34)	7,74 (0,57)
<b>7</b>	4%	19%	5%	35%	8,02 (0,35)	8,37 (0,36)
<b>8</b>	2%	33%	9%	54%	7,78 (0,45)	8,23 (0,39)
<b>9</b>	2%	57%	36%	75%	7,82 (0,46)	8,08 (0,39)
<b>10</b>	2%	32%	12%	57%	7,99 (0,36)	8,04 (0,46)
<b>11</b>	0%	57%	42%	76%	7,94 (0,27)	7,86 (0,44)
<b>12</b>	0%	72%	58%	94%	7,92 (0,43)	8,01 (0,43)



**Figure S1.1:** condition and group effects on response times



**Figure S1.2:** list effects on response times, in interaction with condition (left) and group (right)



**Figure S1.3:** Item position effect on accuracy (up) and response times (down) across group and condition

## Experiment 2

**Table S2.1:** Characteristics of Experiment 2's stories across its two conditions

	<i>Metaphor-base</i>	<i>Synonymic-base</i>	Stat. comparison
Mean length in words (SD)	55.3 (3.9)	54.3 (3.6)	$U = 18, p > .05$
Mean readability (SD) <sup>1</sup>	52.2 (5.9)	49.0 (5.1)	$U = 12.5, p > .05$
Mean content words frequency (SD) <sup>2</sup>	5.7 (0.5)	5.6 (0.6)	$U = 16, p > .05$
Mean referent frequency (SD) <sup>3</sup>	56.5 (7.6)	52.4 (10.7)	$U = 13, p > .05$
Mean target noun frequency (SD) <sup>3</sup>	55.1 (6.3)	48.0 (12.7)	$U = 9, p > .05$

<sup>1</sup> As indexed by the French tool Scolarius ([www.scolarius.com](http://www.scolarius.com)); high score indicate low readability and scores ranging from 50 to 89 being suitable for primary school pupils.

<sup>2</sup> As indexed by the French movie-subtitles database Lexique 2 (New et al., 2004). The median frequency was high, at 297 per million occurrences, this value being at the 97<sup>th</sup> percentile of the whole database. Since frequencies were not distributed normally, we used the log-transformed data to assess difference across conditions.

<sup>3</sup> As measured in the Manulex database (Lété, 2003) gathering French Textbooks for primary schools (1<sup>st</sup> to 5<sup>th</sup> grade). Mean frequency was 54.41 (SD = 9.1) for referents and 51.4 (SD = 10.3) for targets, the overall interquartile range in this database being [24-46].

**Table S2.2:** Estimated TOM effect on accuracy in the different conditions, at different ages, with p-value of the pairwise difference at each age point

Age point	Condition	TOM estimated trend	SE	p value for condition difference
6 years old	Synonymic	-0.33	0.24	< .01
	Metaphoric	0.34	0.16	
7 years old	Synonymic	-0.24	0.18	< .01
	Metaphoric	0.25	0.11	
8 years old	Synonymic	-0.15	0.15	< .05
	Metaphoric	0.15	0.09	
9 years old	Synonymic	-0.06	0.17	.50
	Metaphoric	0.05	0.12	
10 years old	Synonymic	0.03	0.24	.76
	Metaphoric	-0.04	0.17	

**Table S2.3:** Estimated Grammar effect on accuracy in the different conditions, at different ages, with p-value of the pairwise difference at each age point

Age point	Condition	Grammar estimated trend	SE	p value for condition difference
6 years old	Synonymic	0.08	0.07	< .05
	Metaphoric	-0.08	0.05	
7 years old	Synonymic	0.02	0.05	.99
	Metaphoric	0.02	0.04	
8 years old	Synonymic	-0.04	0.06	< .05
	Metaphoric	0.12	0.04	
9 years old	Synonymic	-0.09	0.07	< .001
	Metaphoric	0.21	0.05	

10 years old	Synonymic	-0.15	0.10	< .001
	Metaphoric	0.31	0.06	

**Table S2.4:** Estimated TOM effect on log-transformed response times in the different conditions, at different ages, with p-value of the pairwise difference at each age point

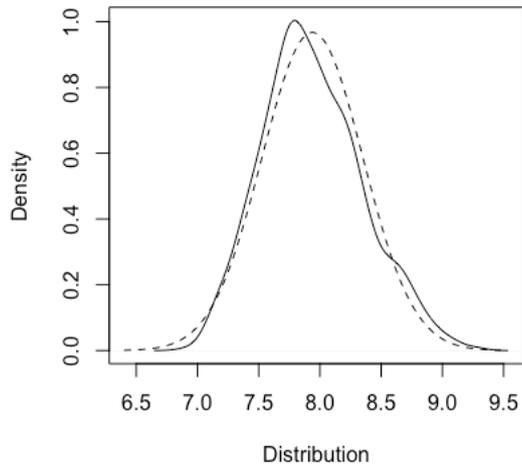
Age point	Condition	TOM estimated trend	SE	p value for condition difference
6 years old	Synonymic	-0.01	0.02	< .05
	Metaphoric	-0.10	0.03	
7 years old	Synonymic	-0.01	0.02	< .05
	Metaphoric	-0.07	0.03	
8 years old	Synonymic	-0.01	0.01	.13
	Metaphoric	-0.04	0.02	
9 years old	Synonymic	-0.005	0.04	.82
	Metaphoric	-0.001	0.02	
10 years old	Synonymic	-0.002	0.03	.20
	Metaphoric	-0.03	0.03	

**Table S2.5:** Estimated Grammar effect on log-transformed response times in the different conditions, at different ages, with p-value of the pairwise difference at each age point

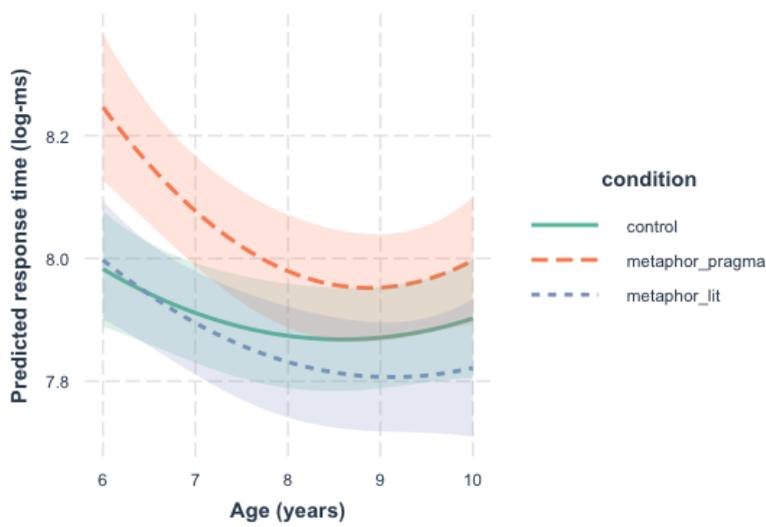
Age point	Condition	Grammar estimated trend	SE	p value for condition difference
6 years old	Synonymic	-0.07	0.008	.41
	Metaphoric	0.002	0.01	
7 years old	Synonymic	0.0002	0.006	.98
	Metaphoric	0.0004	0.008	
8 years old	Synonymic	0.007	0.007	.17
	Metaphoric	-0.002	0.008	
9 years old	Synonymic	0.01	0.007	< .05
	Metaphoric	-0.004	0.008	
10 years old	Synonymic	0.02	0.003	< .01
	Metaphoric	-0.005	0.01	

**Table S2.6:** Estimated TOM effect on accuracy in the different conditions, at different ages, with p-value of the pairwise difference at each age point

Age point	Contrast	Ratio	SE	t-ratio	p value
6 years old	Metaphor-lit vs. metaphor-pragma	1.28	0.064	-0.25	< .0001
	Metaphor-lit vs. synonymic	0.99	0.058	5.02	0.97
7 years old	Metaphor-lit vs. metaphor-pragma	1.20	0.036	6.08	< .0001
	Metaphor-lit vs. synonymic	1.02	0.056	0.30	0.95
8 years old	Metaphor-lit vs. metaphor-pragma	1.16	0.039	4.39	< .0001
	Metaphor-lit vs. synonymic	1.04	0.060	0.75	0.74
9 years old	Metaphor-lit vs. metaphor-pragma	1.16	0.036	4.71	< .0001
	Metaphor-lit vs. synonymic	1.07	0.061	1.13	0.51
10 years old	Metaphor-lit vs. metaphor-pragma	1.19	0.057	3.66	< .001
	Metaphor-lit vs. synonymic	1.08	0.071	1.23	0.45



**Figure S2.1:** Log-transformed response times distribution (plain line) against a normal distribution with the same parameters (dashed line)



**Figure S2.2:** Exploratory analysis. Predicted response times as a function of age for correct responses to literal items, correct (metaphorical) responses to metaphor items and literal responses to metaphor items

## **Annexe D** • Certaines inférences pragmatiques (mais pas toutes) sont difficiles pour les enfants autistes (manuscrit)

*Article en preparation* : Some (but not all) pragmatic inferences are difficult for autistic children: the case of scalar implicatures and metaphors

---

### **Abstract**

Autism is typically associated with difficulties in pragmatic inferences, resulting in an over-literal interpretation of language. This was mostly observed with figurative language, such as metaphors. Yet, the more recent investigation of another type of inferences, that are scalar implicatures, has mostly failed to observe autistic differences, which contrast raises concerns about any general claim of pragmatic difficulties in autism. However, both lines of research face issues: language demands rather than pragmatic competence might actually explain group differences on metaphor tasks, and scalar implicatures were mostly assessed with judgment tasks that might bias their results. The present work thus aims to assess whether this contrast could be observed with tasks that address these concerns, within a single clinical group. We thus employed two innovative tablet tasks, an action-based scalar implicature task, and a picture-selection metaphor task, that both limit language and metalinguistic demand. We assessed a sample of autistic children with no comorbid condition (N = 23), that we compared to a larger sample of typically developing children (N = 237). Surprisingly, compared to the control group, our clinical sample showed a reversed contrast from what was expected, with poorer scalar implicatures comprehension but typical metaphor comprehension, consistently at accuracy and response times levels. We discuss the possibility that a dichotomy of *guided vs spontaneous pragmatic processes*, complementary to previous accounts, might explain this result. Paralleling theory of mind hypothesis, autistic individuals could show typical pragmatic abilities, as measured with guiding paradigms, but might experience differences in the spontaneous use of those skills.

---

Autism spectrum disorder (from now on: autism) is a heterogeneous neurodevelopmental condition characterized by difficulties in communication and social interactions, and atypical or restricted behaviors, activities or

interests (American Psychiatric Association, 2013). Autism is frequently associated with other neurodevelopmental conditions, especially language disorders or intellectual development disorders (Levy et al., 2010). However, even when it is not the

case, autistic children<sup>27</sup> experience communication issues, among which they are typically thought to struggle with pragmatic inferences, resulting in an excessively literal interpretation of language. This hallmark of autism crossed the ages from the early descriptions of Kanner (1946) or Asperger (1944) to the current diagnostic criteria (APA, 2013), being also reported by autistic individuals themselves (Müller et al., 2008; Wilson, 2022) and having infiltrated public representations of autism through mainstream culture (Semino, 2014).

This difference is classically observed with figurative language such as metaphors, starting from Happé's (1993) seminal study which has been replicated in several studies (e.g., Rundblad & Annaz, 2010), and as shown in recent meta-analyses in children and adults with both accuracy measures and response times (Kalandadze et al., 2018, 2019; Morsanyi et al., 2020). Effect sizes, however, largely vary and are influenced by group-matching strategies and task properties (Kalandadze et al., 2018; Morsanyi et al., 2020). Verbal explanation tasks typically yield larger effect sizes but are criticized for their metalinguistic demand (Pouscoulous, 2011), which confounds whether group-differences can be attributed to non-autistic differences such as variation in structural language skills (see Norbury, 2005; Kalandadze et al., 2019).

Moreover, this historical claim (see also Surian, 1996) has been challenged by the more recent investigation of other types of pragmatic inferences in autism, such as scalar implicatures. Scalar implicatures designate uses of terms embed in a scale, such that a weaker term implicates the

negation of a more informative term (Horn, 1989). The most stereotypical example is the use of *some* as the denial of the more informative *all*, although *some* is semantically compatible with *all*. Hence, when presented with underinformative sentences, adults (more than children) are typically inclined to reject those statements. Since the enrichment of *some* as *meaning some-but-not-all* is arguably pragmatic in nature, and not logical, autistic individuals were expected to present differences with their neurotypical pairs. Yet, almost all attempts failed to observe such effect, be it in children, adolescents and adults (Andrés-Roqueta & Katsos, 2020; Chevallier et al., 2010; Hochstein et al., 2017; Pijnacker et al., 2009; Schaeken et al., 2018; Su & Su, 2015). Importantly, these studies employed sentence judgement paradigms, by asking participants to judge whether they agreed with underinformative utterances. These tasks, however, should not be considered as direct measure of implicatures derivations, as alternative pathways can lead to rejecting underinformative utterances (see Katsos & Bishop, 2011; Katsos, 2014; or Kissine & De Brabanter, 2023), but rather as a proxy, or as an indirect reflection of scalar implicatures. Interestingly, one of these studies also employed a so-called ternary-judgement task (Schaeken et al., 2018), in which children were given the possibility of an intermediate judgement between “agree” and “disagree” (“I agree a bit”). This revealed a clear group difference: typically developing children largely preferred this intermediate response for underinformative statements, while autistic children were more likely to reject completely such sentences. Two other studies used multiple-choice tasks and tended to observe a difference between autistic and

---

<sup>27</sup> Regarding the way to refer to autistic individuals, which has important real-life consequences (Vivanti, 2020), in absence of a clear consensus of the preferred terms (although

a consensus emerges on the least preferred), we followed Botha et al.'s (2021) rationale and used identity-first formulation, considered as less offensive and less stigmatizing.

neurotypical children (Mazzaggio et al., 2021; Pastor-Cerezuela et al., 2018), even though methodological issues raise caution on the interpretation of these results. First, Pastor-Cerezuela et al. (2018) used a written multiple-choice task, where children were asked to select the best interpretation of underinformative sentences. Autistic children performed poorer than age or language-matched pairs. However, the task mixed together scalar and non-scalar inferences, which does not allow for conclusion on scalars specifically, and it did not include control items, which raises concerns about whether autistic children were struggling with underinformative statements specifically, or with the task in itself. Second, Mazzaggio et al. (2021) used a picture-based multiple-choice task, with orally presented material, with 4 to 9 years old children. Autistic children seem to perform poorer in scalar implicatures in the exploratory analyses, but the main analysis failed to detect the critical condition by group interaction, indicating that ASD children performed poorer in general, but not specifically in underinformative sentences, also on control items. Whether autistic individuals perform typically on scalar implicature is thus still to be confirmed with a different paradigm than truth judgment tasks only.

The contrast between some poorer and some typical pragmatic inferences in autism, compared to neurotypical individuals, contributed to the fragmentation of pragmatics as a unitary competence, and led some researchers to propose a dichotomy opposing linguistic- vs social-pragmatics (Andrés-Roqueta & Katsos, 2017, 2020). Linguistic-pragmatics designate pragmatic operations which depend on linguistic competence and pragmatic knowledge, while social-pragmatics additionally require reasoning on the speaker's mental states, that is,

theory of mind (TOM, Premack & Woodruff, 1978). In this view, autistic individuals with normal linguistic skills are expected to perform similarly to neurotypical controls on linguistic-pragmatics, but poorer on social-pragmatics, given the assumed autistic difference in TOM processes (Happé et al., 2017). Scalars implicatures have thus been exemplified as linguistic-pragmatics, leaning on the fact that verbal abilities were shown to be a good predictor of such abilities (Andrés-Roqueta & Katsos, 2020; Chevallier et al., 2010; Pijnacker et al., 2009). In contrast, figurative language is rather considered an instance of social-pragmatics (Andrés-Roqueta & Katsos, 2020). However, while this distinction emerged from the comparison of different studies, with different sample, comparisons of different types of pragmatic inference within a single autistic sample exist but are much scarcer (Andrés-Roqueta & Katsos, 2020; Deliens et al., 2018).

In this context, the aim of our study is three-fold. First, we wanted to assess whether autistic children's typical performance with scalar implicatures could be replicated with a naturalistic task which does not tap on explicit metalinguistic processes. Second, we wanted to confirm that autism was associated with difficulties in processing metaphors, in a new referential metaphor task which also limits language and metalinguistic demands. Third, we aimed to test whether these tasks confirmed the contrast between scalar implicatures and metaphor comprehension, as an example of social-pragmatics, in a single clinical sample. School-age is arguably an appropriate window to observe such effects, since both scalar implicature and metaphors undergo important improvement at that age (e.g., Deckert et al., 2019; Katsos et al., 2016; Winner et al., 1976). Based on the literature, we hypothesized that autistic children with typical language abilities

would perform similarly to their neurotypical peers on scalar implicatures, while performing poorer on metaphors.

## Methods

### Participants

Twenty-three autistic children (3 girls) were recruited at the hospital *Le Vinatier* for this study, aged from 6;7 to 10;8 years old. All of them had received an ASD diagnostic by trained child psychiatrists and met the DSM-5 diagnostic criteria, which was confirmed by gold-standard measures extracted from their medical files (ADOS-2, SCQ and/or ADI-R), as well as the exclusion of intellectual development disorder. We report these measures for each participant in supplementary materials (table S1). Regarding intellectual abilities, recent measures were not always available (missing = 3), and were collected with various instruments (see detailed table in supplementary material), but full-scale IQ was in the normal to normal-superior range (mean = 111, 95% CI = [100,122], range = [79, 138], missing = 10), with verbal functioning in the normal-superior range (mean = 107, 95% CI = [101, 113], range = [78, 130], missing = 4) as well as fluid reasoning (mean = 113, 95% CI = [106, 120], range = [82, 135], missing = 4). Comorbid attention disorder or language disorder were considered as exclusion criteria, to avoid potential confounds. All of the children attended regular schools and were native French speakers.

To provide control data, we used a set of 237 children from previous studies (*Petit et al., chapt. 4; Petit et al., chapt. 5*), recruited in two French schools, aged between 6 and 11 years old, who were native French speakers and whose parents reported no neurodevelopmental disorder. To make sure that this group represented typical

development, it only included children whose performance were not below 2.5 SD under their age groups' mean on a receptive grammar task and a theory of mind task (see *Petit et al. experiment 2*).

All children's socioeconomic status (SES) were measured through parental questionnaires of parents' education level (coded from 0 = no diploma to 7 = PhD) and family resources via the Family Affluence Scale (Currie et al., 2008), standardized in a 0 to 100 score (from the minimum to the maximum possible value), and then averaged in a SES composite, as had been done for the control group.

All participants and their parents provided informed consent. This study was part of a larger project on pragmatic inferences across development, which received ethical approval from the local IRB (Comité de Protection des Personnes Sud-Est I, ID RCB 2019-A01721-56).

### Material

Children were tested with two French validated language tests from the BILO-3C (Khomsi et al., 2007), to confirm the exclusion of language disorder and better characterize the receptive linguistic abilities of autistic participants. More specifically, we assessed receptive grammar (CO subtask) through a computerized image matching task resulting in a raw score ranging from 0 to 27 and receptive vocabulary (JL subtask) through a lexical judgement task resulting in a score ranging from 0 to 43.

We also used 2 new clinical tasks from the French tablet-based battery TIPi, assessing scalar inferences (*Petit et al., chapt. 4*) and novel metaphor comprehension (*Petit et al., chapt. 5*). Both tasks are designed to limit language and metalinguistic demands, so as to limit possible confounds in the assessment. The *scalar task* is an action-based paradigm in which a table with 5

plates and fruits are displayed on the screen; children are asked to help set the scene, following oral instructions such as “*I want all plates to have a banana*” (originally in French, “*Je veux que toutes les assiettes aient une banane*”). Participants can afterward modify the setting by sliding fruits on the screen, and then are to validate when they are satisfied with the disposition. Variations are applied to initial fruits disposition (all plates empty, all plates filled, only a subset of the plates filled) as well as the instruction’s quantifier (none, some, all<sup>28</sup>), so as to create all possible conditions, including a critical condition, in which *all* plates initially contain a fruit, but the instruction is that *some of the plates* should. Here, a semantic interpretation of the scalar term (‘some-and-possibly-all’) would directly lead to validation, while a pragmatic interpretation (‘some-but-not-all’) would need at least one plate to be cleared before validation. This critical underinformative condition represents 7 items, mixed with the other conditions and filler items (where different fruits are used in the same item, to limit repetitiveness) in a pool of 29 total items (see detailed description in *Petit et al., chapt. 4*). Children’s responses at validation as well as response times are recorded (response times include the duration between the disambiguating phoneme of the vocal instruction and the validation by the participant). Critical items are automatically categorized as pragmatic or semantic responses. All items, including controls, are additionally coded as semantically erroneous (when the response violates the quantifier’s semantic meaning) or not. The task is preceded with 4 training items, using literal content only, which

must be succeeded for the proper task to start. The task lasts about 5 minutes.

The *novel metaphor comprehension task* is a referential task, in which oral verbal stories are read to children. A referent is introduced in the first part of the story, and then recalled by a character in the final critical sentence. This reference can be either metaphorical, in the critical condition, or synonymic in a control condition. For example, consider the expression *the fountain*, which harks back to a previously-mentioned element; in the metaphorical-reference condition, the expression refers to a crying baby and in the synonymic-reference condition it refers back to water jets in a park that need repairing (see the structure of the task in figure S1 of the supplementary materials). Metaphors were designed to be novel. A question on the nature of referent is asked to the child afterwards, to which they are expected to answer by selecting one picture out of four, including a metaphorical interpretation, a literal interpretation, a distractor and an I-don’t-know response option. In the case of synonymic reference, the metaphorical interpretation becomes a simple distractor, while in the metaphorical condition two response options are typically attracting to children: the correct (intended) figurative interpretation of the referent, and the literal representation of the reference. Questions target what happened concretely in the story, needing the reference to be understood, rather than focusing on the *intention* of the character. Stories are 47 to 63 words long, and the vocabulary is kept simple, especially for the target and the reference nouns. While stories are uttered, a unique picture is displayed on the tablet, representing its general context, in order to favor attention engagement (rather than

---

<sup>28</sup> English translation presented, the original French quantifiers were, respectively, “aucune” (none), “certaines” (some), “toutes” (all)

comprehension of the story *per se*). Responses as well as responses times are recorded. The task includes 6 novel metaphors and 6 literal stories. In order to mask the task's goal and structure to children, the 12 items are mixed with other stories involving other kind of linguistic or pragmatic phenomena that are not of interest to this paper and considered as fillers. The task is preceded with 3 training items, using literal content only, which must be succeeded for the proper task to start, and is about 20 minutes long.

### Procedure

Children from the autistic group were tested individually in a quiet office of the hospital. Tasks order was randomized and adapted to each participant's motivation and availability. For example, breaks were proposed when children showed signs of attentional fatigue, longer tasks were proposed when children were thought to be maximally available, while more motivating tasks could be proposed when children manifested signs of boredom. Typically developing children had been tested in collective settings (in their classrooms), with individual tablets and headsets (see *Petit et al., chapt. 4*, for precise description). With this group the metaphor task, which is longer, was proposed first, followed by the scalar task.

### Analysis

Analyses were run on R (R Core Team, 2022). Accuracy was analyzed following Thomas et al.'s developmental trajectory approach, which recommends regressing performance against age for each group and comparing the trajectories, in order to provide richer information than a simple mean comparison; this methodology takes advantage of the samples age-diversity, rather than being hindered by it (Thomas et al., 2009). We first applied it as a control analysis for the language tests, with

multiple linear regressions with total performance as the dependent variable and age, group and their interaction, as predictors. Then we turned to our experimental tasks, for which we used generalized mixed effect models with age and group, and their potential interaction, as fixed effect, and random intercepts for participants and items. For the scalar task, two separate models were run, one explaining semantic errors on all items, as a control analysis, and one explaining pragmatic responses on critical items. For the metaphor task, condition (metaphor vs. synonymic) was added as a fixed effect, together with its potential interactions with the other effects. Since gender showed no effect on performance in the validation studies of our experimental tasks (*Petit et al., chapt. 4; Petit et al., chapt. 5*), it was not included in the analyses.

Response times were then analyzed with mixed effect models, after log-transforming the data to correct for typical skewness, excluding incorrect responses and removing outliers, defined as responses which deviated of more than 2.5 DS from groups and conditions' means. For metaphors, we specified group and reference-condition as fixed effects, with their interaction, and by-item and by-participant random intercepts. The scalar task imposed a more complicated methods, since we wanted to compare pragmatic responses (where fruits had to be moved) to semantic responses to critical items (where no fruits had to be moved), and to responses to control situations (responses to control quantifiers, where both situations happened). Yet response times include the time needed to move fruits, which needed to be controlled for in order to isolate the effect of our variable of interest. To do so, we followed Petit et al. (*chapt. 4*)'s methodology and added to the previous effects whether participants directly validated the situation or not (*hasMoved*,

binary variable) and, in the latter case, how many fruits were moved (*moves*, continuous variable). Both were specified as fixed effects and as random slopes for participants for the most conservative control.

Models were fit with the *lme4* package (Bates et al., 2015) and contrasts estimated with *emmeans* (Lenth, 2022), adjusting for multiple comparison with Tukey method when necessary. Dummy coding was used for group (TD as baseline) and sum contrasts for the other contrasts, especially condition.

## Results

### Control language measures

Background measures and descriptive performances of both groups on the different tasks are displayed in table D-1. We first analyzed performance on the control language tasks (see figure S1 in supplementary materials). For the receptive grammar task, the two developmental trajectories were remarkably superimposed, revealing a main effect of age ( $\beta = 1.36$ ,  $SE = 0.16$ ,  $p < .001$ ), but no main effect of

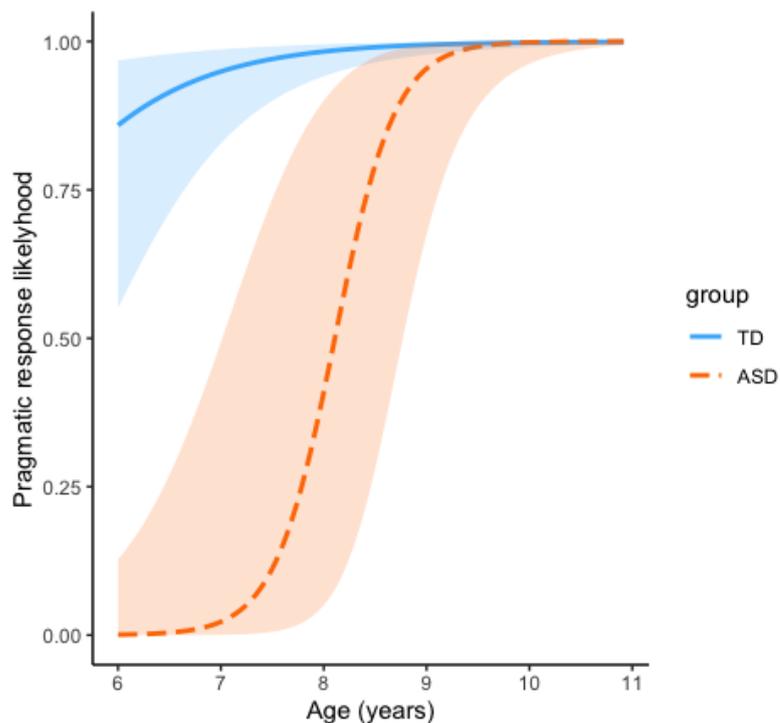
group ( $\beta = 1.8$ ,  $SE = 4.8$ ,  $p = .70$ ) and no interaction ( $\beta = -0.20$ ,  $SE = 0.54$ ,  $p = .71$ ). For the receptive vocabulary task, there was a main effect of age ( $\beta = 1.3$ ,  $SE = 0.17$ ,  $p < .001$ ) and of group ( $\beta = 10.8$ ,  $SE = 5.35$ ,  $p < .05$ ) with an almost significant interaction term ( $\beta = -1.15$ ,  $SE = 0.6$ ,  $p < .10$ ): ASD children tended to outperform TD children in the younger section of the sample.

### Scalar implicatures

To analyze accuracy in the scalar task, we first looked at semantic errors, which were very rare (2% of the responses). The model showed that they were slightly less likely to occur with higher age (OR = 0.67, 95% CI = [0.48, 0.93],  $p < .05$ ) but there was no effect of group (OR = 0.58, 95% CI = [0.15-2.22],  $p > .05$ ) and no interaction (OR = 2.39, 95% CI = [0.60-9.48],  $p > .05$ ). As this confirmed that the task was accessible to children, independently of group, we turned to our critical condition, where pragmatic responses were more variable (77% of responses overall). The likelihood of pragmatic responses was shown by the model to be influenced by age (OR = 4.87, 95% CI = [2.4, 9.9]) and by group (OR = 0.03, 95% CI = [0.00, 0.38],  $p < .01$ ), with

**Table D-1:** Samples' characteristics and descriptive statistics for control and experimental tasks

	N	237	23	
Demographic information	Girls	51%	13%	$\chi^2(1, N = 260) = 0.00, p < .05$
	Mean age (SD)	8.4 (1.4)	8.7 (1.4)	$t(258) = 0.98, p = .33$
	Mean SES (SD)	58.8 (15.7)	53.2 (17.0)	$t(258) = 1.62, p = .11$
Language control tasks	Mean receptive vocabulary (SD)	34.7 (4.2)	35.8 (3.2)	
	Mean receptive grammar (SD)	15.9 (3.8)	16.5 (3.9)	
Metaphor task	Mean accuracy, literal condition (SD)	95% (12%)	94% (11%)	
	Mean accuracy, metaphor condition (SD)	44% (35%)	49% (37%)	
Scalar task	Mean error rate (SD)	2% (0.6%)	2% (0.5%)	
	Mean pragmatic response rate (SD)	79% (34%)	65% (43%)	



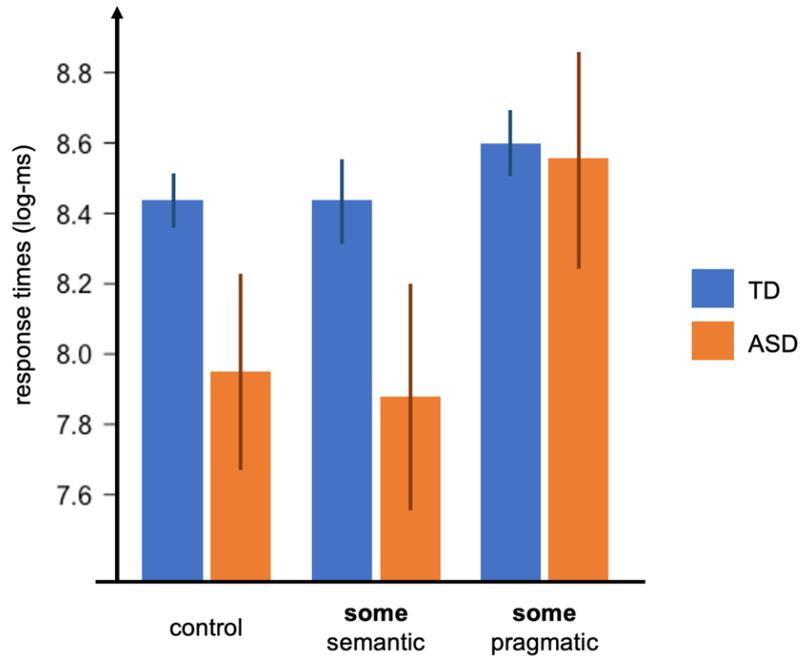
**Figure D-1:** predicted probability of pragmatic responses in the scalar task based on age and group

a group by age interaction ( $OR = 24.7$ , 95%  $CI = [2.0, 297.7]$ ,  $p < .05$ ): pragmatic responses were clearly less likely to occur in the ASD group, and this difference was more pronounced in the younger portion of the sample (see figure D-1). We summarized for each participant their pragmatic interpretation rate for the critical condition, which reveal bimodal distributions in both groups, with a peak of consistently pragmatic responders, and a peak of consistently semantic responders (see figure S2 in supplementary materials).

As an exploratory analysis, we assess children's tendency to operate changes on the scenes, independently from the instructions. To do so, we isolated control items for which changes were possible but not necessary, that is, 'some' items with a partial initial context (e.g., when the instruction was "I want some plates to have a banana" when 2 plates initially did: even if this was not necessary, children still had the possibility to add 1 or 2 bananas without changing the answer's categorization). We

excluded errors. Then we fitted a generalized model explaining the probability to make a change, with age, group and their interaction as fixed effects, as well as random intercepts for participants and items. This revealed that the probability to operate unnecessary changes was important but reliably decreased with age ( $\beta = -0.54$ ,  $SE = 0.12$ ,  $z = -4.3$ ,  $p < .001$ ), and was more important in the autistic group ( $\beta = 1.3$ ,  $SE = 0.45$ ,  $z = 2.9$ ,  $p < .01$ ). No interaction was detected ( $p = .79$ ).

Turning to response times, we were interested in the effect of response-type (semantic responses to the critical condition vs pragmatic responses to that condition vs responses to control quantifiers, 3637 data points), together with the control variables described in the methods section. This revealed a group by response-type interaction (likelihood ratio test  $X^2(2) = 41$ ,  $p < .0001$ , complete model's output in table D-2). Post-hoc contrasts revealed the same pattern in both groups (figure D-2): pragmatic responses were slower than both



**Figure D-2:** Response-type and group partial effects and interaction on response times in the scalar task; error bars represent 95% CI.

**Table D-2:** Outcome of the model explaining response times in the scalar task

	LOG(RT)		
PREDICTORS	Estimates	95% CI	p
(INTERCEPT)	8.02	7.95 - 8.09	<0.001
HASMOVED	0.06	-0.10 - 0.22	0.435
MOVES	0.22	0.20 - 0.25	<0.001
RESPONSE-TYPE1	-0.17	-0.37 - 0.03	0.102
RESPONSE -TYPE2	-0.16	-0.40 - 0.07	0.173
GROUP [ASD]	<b>-0.37</b>	<b>-0.57 - -0.16</b>	<b>&lt;0.001</b>
RESPONSE -TYPE1 * GROUP [ASD]	-0.36	-1.03 - 0.31	0.296
RESPONSE -TYPE2 * GROUP [ASD]	<b>-0.58</b>	<b>-1.00 - -0.16</b>	<b>0.007</b>

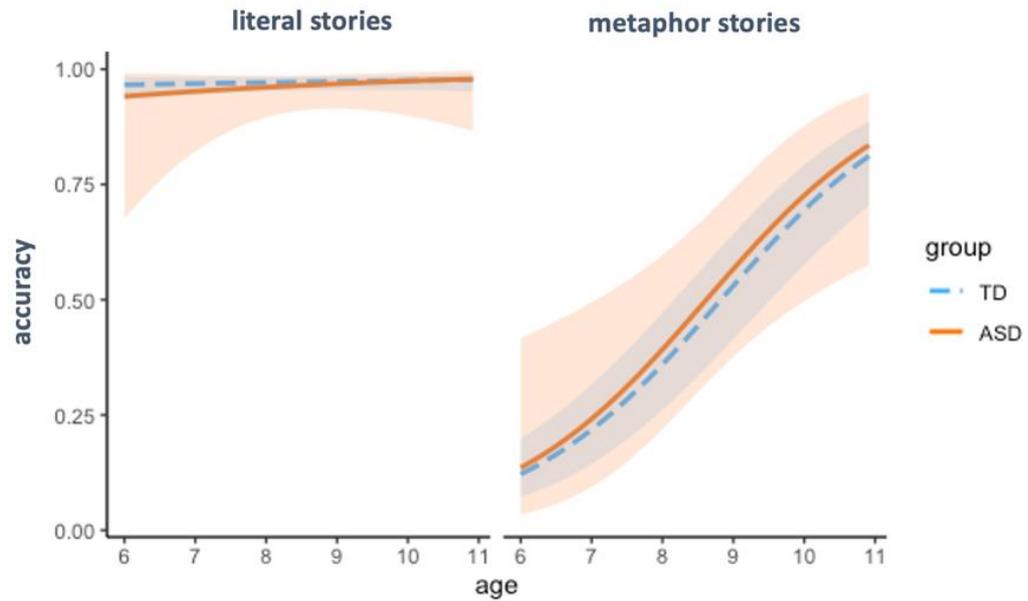
semantic responses and control responses (all  $p$ s < .05), but this difference was amplified in the ASD group, which was clearly faster on both semantic responses ( $\beta = 0.56$ ,  $SE = 0.14$ ,  $p < .001$ ) and control responses ( $\beta = 0.48$ ,  $SE = 0.13$ ,  $p < .001$ ), but was as slow as TD participants when responding pragmatically ( $\beta = 0.05$ ,  $SE = 0.14$ ,  $p > .05$ ).

### Metaphors

In the metaphor task, accuracy was very high in the synonymic control condition (95%) and mixed in the metaphor condition (45%), where most of the errors resulted from the selection of the literal response

option (94% of the errors). The model (table D-3) revealed a clear age by condition interaction but, crucially, no group effect or group interaction with other factors. In other terms, the developmental trajectories of both groups were remarkably similar on both conditions (as visible on figure D-3), with literal stories at ceiling as soon as 6 years old, but an important progression on metaphors with increasing age. Exploratory by-item analysis revealed that this pattern applied to each item taken individually.

Regarding response times, the model was run on 2125 data points, after the exclusion of 2% of outliers. It revealed main effects of condition ( $\beta = -1.16$ ,  $SE = 0.06$ ,  $p < .05$ ) and



**Figure D-3:** predicted accuracy in the metaphor task for both groups based on age and condition

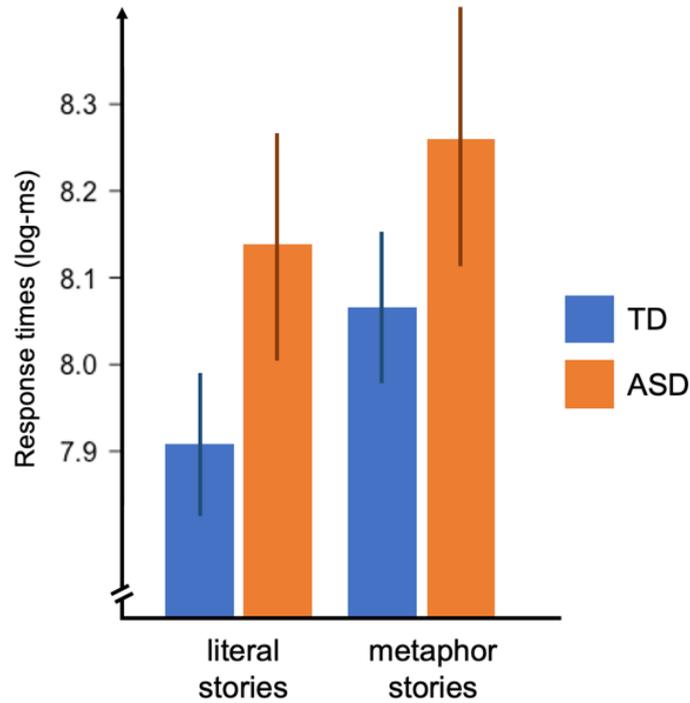
**Table D-3:** Output from the generalized model explaining accuracy in the metaphor task

Accuracy			
Predictors	Odds Ratios	95% CI	p
(INTERCEPT)	5.01	3.48 - 7.20	<0.001
<b>CONDITION</b>	<b>46.75</b>	<b>24.14 - 90.52</b>	<b>&lt;0.001</b>
<b>AGE</b>	<b>1.76</b>	<b>1.43 - 2.18</b>	<b>&lt;0.001</b>
<b>GROUP</b>	0.87	0.43 - 1.75	0.691
<b>CONDITION * AGE</b>	<b>0.43</b>	<b>0.32 - 0.58</b>	<b>&lt;0.001</b>
<b>CONDITION * GROUP</b>	0.70	0.28 - 1.76	0.443
<b>AGE * GROUP</b>	1.21	0.57 - 2.56	0.622
<b>(CONDITION * AGE) * GROUP</b>	1.18	0.42 - 3.29	0.754

group ( $\beta = 0.21$ ,  $SE = 0.05$ ,  $p < .001$ ), but no interaction ( $\beta = 0.03$ ,  $SE = 0.06$ ,  $p > .05$ ). Metaphors prompted longer response times, and ASD participants were slower in general (see figure D-4). In sum, the temporal cost associated with the metaphoric condition, compared to the control synonymic condition, was of similar size in both groups.

In order to get a finer appreciation of how the participants performed the task, we exploratorily investigated the status of the(over)literal interpretations of metaphoric references, which were considered as errors and excluded from the main analysis. To do so, we added these responses to the previous data set, creating a new response-type variable opposing

response to synonymic items vs correct responses to metaphor items vs. literal responses to metaphor items (2972 data points, detailed output in table D-4). Here, a group by response-type interaction appeared, revealing that literal interpretation of metaphors had a different status across groups. In the TD group, literal interpretations of metaphors appeared faster than correct figurative interpretations ( $\beta = -0.16$ ,  $SE = 0.02$ ,  $p < .0001$ ) but as fast as response to literal items ( $\beta = 0.01$ ,  $SE = 0.06$ ,  $p > .05$ ). It was the opposite in the ASD group: literal interpretations of metaphors were as long as correct figurative interpretations ( $\beta = 0.10$ ,  $SE = 0.08$ ,  $p > .05$ ), but slower than responses to literal items ( $\beta = -0.23$ ,  $SE = 0.08$ ,  $p < .05$ ). That is, literal interpretations of metaphors were



**Figure D-4:** predicted response times in the metaphor task for both groups in both conditions

**Table D-4:** Output from the exploratory model explaining response times in the metaphor reference task after the inclusion of literal responses to metaphors (response to literal items vs. literal responses to metaphor items vs correct responses to metaphor items)

PREDICTORS	LOG(RT)		
	Estimates	95% CI	p
(INTERCEPT)	7.96	7.89 - 8.02	<b>&lt;0.001</b>
RESPONSE-TYPE1	-0.14	-0.36 - 0.08	0.209
RESPONSE-TYPE2	-0.17	-0.30 - -0.04	<b>0.010</b>
GROUP	0.31	0.21 - 0.42	<b>&lt;0.001</b>
RESPONSE-TYPE1* GROUP	-0.22	-0.42 - -0.03	<b>0.027</b>
RESPONSE-TYPE2* GROUP	0.49	0.24 - 0.75	<b>&lt;0.001</b>

comparable to synonymic items in the TD group, but similar to correct responses to metaphors in the ASD group.

## Discussion

In this study, we aimed to test whether autistic children’s typical performance in scalar implicatures could be observed with a naturalistic task that does not tap on explicit metalinguistic processes, whether autistic children had difficulties with a new

metaphor referential task, and whether such contrast could be observed within a single sample. We thus tested with recent tasks 23 autistic children and compared their performances to 237 typically developing children from previous studies. Conversely to what we hypothesized, we observed that autistic children performed poorer than neurotypical peers in scalar implicatures but similarly on metaphors comprehension. Response times patterns were consistent with this view, since the autistic children showed an increased cost of pragmatic

responses compared to control items in the scalar task, but a similar temporal cost associated with metaphors. We first discuss how these results fit with both the literatures on scalar implicatures and metaphors, then examine why this contrast was unexpected; we finally consider the need for an alternative explanation.

### **Scalar implicatures**

Our results thus contrast with most of the existing studies so far, which observed that autistic participants processed scalar implicatures typically (Andrés-Roqueta & Katsos, 2020; Chevallier et al., 2010; Hochstein et al., 2017; Pijnacker et al., 2009; Schaeken et al., 2018; Su & Su, 2015). Exploratory analysis suggested that the group difference in our results could not be attributed to peculiarities in the management of the task itself, rather than with scalar implicatures, since, 1) no group difference could be observed on control items, and 2) autistic children were more likely to make *unnecessary changes*, which in the critical condition would lead to more responses categorized as pragmatic, while we observed the very opposite. Moreover, the characteristics of our sample rule out the possibility that this pattern could be attributed to other, irrelevant differences than the diagnosis, since the autistic children we tested presented good to superior formal language and intelligence skills, and performed similarly to controls in the control analysis. Groups only differed in gender, but we previously showed that this variable had no effect on our tasks (*Petit et al., chapt. 4; Petit et al., chapt. 5*). Taken together, our results and previous studies thus suggest that autistic individuals are typically able to judge the appropriateness of a pragmatic utterance *when asked to* (although atypically when given the possibility of an intermediate judgement, see Schaeken et al., 2018), but that they are less likely to *spontaneously* use

informativity to derive speakers' intentions, resulting in fewer pragmatic interpretations, and at a greater cost in terms of response times, relatively to control conditions. Accordingly, studies which did not employ binary judgement but multiple-choice response format tended to observe a diminished rate of pragmatic responses in autistic children (Mazzaggio et al., 2021; Pastor-Cerezuela et al., 2018).

### **Metaphor comprehension**

Our results also contrast with previous evidence which reported an impaired comprehension of metaphors in autistic compared to neurotypical individuals (Kalandadze et al., 2018, 2019; Morsanyi et al., 2020), while we observed very similar developmental trends in our groups. One might consider possible that such effect failed to reach significance due to the limited size of our autistic sample, but we consider this possibility unlikely, for three reasons. First, previous studies often used similar sample size (see for instance the studies included in Morsanyi et al.'s 2020 meta-analysis). Second, the specific task we used showed a good sensitivity to interindividual differences in previous work (*Petit et al., chapt. 5*) as well as in this study, as revealed by the clear developmental pattern observed in the critical condition (figure D-3). Third, this specific autistic group showed an atypical developmental curve in the other task we used and that we just discussed. Previous evidence, however, showed considerable variation in effect sizes which has been related, notably, to group matching strategies (Kalandadze et al., 2018) and experimental variables (Kalandadze et al., 2019; Morsanyi et al., 2020): effects sizes were typically lower when groups were matched on language abilities and when using multiple-choice instead of verbal explanations. Our study matches both of those criteria, which make its results less

surprising, even though we did not expect an absence of group effect. A possible explanatory factor is that, although our groups were matched in terms of age and language abilities, the autistic children we tested had a slight advantage in terms of non-verbal intelligence, which might have hidden subtle group differences. Recent evidence, however, showed that the link between metaphor comprehension and analogical reasoning of autistic participants disappeared once formal language skills (vocabulary) was taken into consideration (Morsanyi et al., 2022). Another possible factor lies within the task we used, which was conceived to limit linguistic demand (see *Petit et al., chapt. 5*), not only in the response format but also in the linguistic material itself, with frequent vocabulary and relatively simple syntactic structures, as confirmed by the ceiling performances of both groups in the literal condition. Anyway, our results confirm that autistic individual can interpret metaphors typically, in those conditions.

Regarding how these metaphors were processed, response times can be used to grasp underlying mechanisms. Few previous studies have used this kind of measures with autistic participants, and reported group differences, even in cases where accuracy was comparable between groups (Chahboun et al., 2016; Morsanyi et al., 2022). However, while those few studies reported slower responses to metaphorical material in autistic compared to neurotypical participants, they did not include control material involving literal meaning only, raising doubts about whether autistic participants were slower with metaphors, or with these tasks in general. In our study, autistic participants did respond slower to metaphors, but so did they with literal control items, with no interaction between group and condition. In other words, the temporal cost associated with metaphors comprehension, compared to

literal comprehension, was similar in both groups. This suggests atypical processes involved by the task in general, rather than with metaphorical expressions specifically, which could also explain the results of previous studies. Accordingly, the other studies with autistic adults which included control conditions did not observe clear group by condition interactions (Gold et al., 2010; Gold & Faust, 2010).

The only atypical pattern associated with autism appeared in the exploratory analysis focused on the status of over-literal interpretations of metaphors. In the TD group these errors were provided as fast as responses to the control condition, suggesting that these errors reflect a lack of pragmatic reasoning. In contrast, autistic children showed a similar slowdown when they interpreted metaphors literally or figuratively. It might be that even when they responded literally to metaphor, the metaphorical interpretation remained a possible option, conversely to TD children. Indeed, some studies suggested that the generation of metaphorical meaning would be normal in autistic participants, and that possible difficulties would lie in the subsequent selection phase between meaning competitors (Chouinard & Cummine, 2016; Hermann et al., 2013). This hypothesis, however, deserves further investigations, especially since those results were exploratory.

### **Contrasting scalars and metaphors**

After discussing how the results we obtained with each of our task fitted in the literature, we can turn to our more global research question, which was whether a contrast between scalar implicatures and metaphor comprehension could be observed within a single autistic group. We recall that our study controlled for language confounds by (i) selecting groups which were matched in terms of language abilities and (ii) using two tasks which limit

linguistic demand and propose non-verbal response formats. Based on the linguistic-vs social-pragmatics distinction (Andrés-Roqueta & Katsos, 2017), which stands that autistic individuals' performance with scalar phenomenon would depend on the TOM demand of the situation, we thus expected our autistic participants to succeed the linguistic-pragmatics task (scalar implicatures) but to perform poorer in the social-pragmatics task (metaphors). And yet, as discussed above, we did observe a contrast, but in the opposite direction. While autistic children showed scarcer and more effortful scalar implicatures than typically developing controls, they demonstrated a typical comprehension of metaphors. We first acknowledge, following Marocchini (2023), that this is not because autistic participants succeed in a task that this task does not involve social reasoning or, more specifically, theory of mind, despite the heritage of the so-called *mental-blindness* theory of autism. Indeed, we previously showed that the particular metaphor task we used was found to be associated with theory of mind performance, especially in younger children (*Petit et al., chapt. 5*). We thus argue that the contrast we observed calls for another dichotomy.

### **A need for a novel dichotomy?**

As discussed in the metaphor literature, lowering linguistic and interaction related demands is typically considered as a facilitating factor for autistic individuals (Kalandadze et al., 2019; Morsanyi et al., 2020). Yet, the scalar task and the metaphor task we used are both designed to limit those demands; so this aspect cannot explain the contrast we observed. On the other hand, another factor can distinguish the two tasks we employed: the scalar task demands the *spontaneous* use of information to derive the speaker's intention, while the metaphor task *guides*

the pragmatic reasoning, with questions and response propositions. This thus suggests that autistic individuals could show typical pragmatic abilities, as measured with guiding paradigms, but that they could struggle to use those skills spontaneously, resulting in poorer performances in task which do not guide pragmatic reasoning, and in difficulties in everyday life interactions with (mostly non-autistic) speakers.

We argue that this distinction might contribute to explain the stronger effect sizes observed in metaphor studies which employed verbal explanation tasks, which is currently attributed to language demands only: these tasks arguably guide less pragmatic reasoning compared to judgement tasks or multiple-choice response format. For example, language demand was used to explain why Rundblad & Annaz' (2010) study yielded the largest effect size in Kalandadze et al. (2019) meta-analysis. Interestingly, this study used a task that is in some respects similar to ours, as it uses verbal stories ending with a metaphor, followed by a question on the referent of this metaphor. As in our task, questions targeted concrete elements and avoided mental terms (for example "What does Stuart see?" after a crowd was described as a "flood"). However, an important distinction between our paradigms is that we proposed response options, while children had to respond orally in Rundblad & Annaz (2010). Although Kalandadze et al. (2019) classified this task as an "open verbal explanation task" (p. 1441), we argue that children were not required to *explain* the meaning of the metaphors, and that a single-word response was enough. Hence, the language demand of this task does not appear as much higher than in our task, which yielded very different results (for similar results with novel metaphors, see Van Herwegen & Rundblad [2018]).

However, we argue that open questions guide much less pragmatic reasoning than multiple choice, and that this factor could play an important role in explaining the large discrepancy between our results. Previous studies have directly showed that subtle modifications in experimental settings could have large impacts on group-differences in autism studies. Wilson & Bishop (2021) for example showed that when presented with questions on implied meanings, autistic adults were much more likely to select an “I don’t know” response when available. But when this proposition was removed, they responded similarly to non-autistic adults in most cases. This revealed that autistic participants’ management of pragmatic inferences was somewhat related to a general preference for certainty, and that response propositions could *guide* pragmatic reasoning and force participants to reach a conclusion, while they might spontaneously remain on an uncertain outcome.

Conversely, compared to classical judgement paradigms, our action-based scalar task reduces metalinguistic demands and boosts typically developing children’s performances (*Petit et al., chapt. 4, Pouscoulous et al., 2007*). Yet, while it should also facilitate testing with autistic children, it showed the opposing effect on them: their performances were relatively hindered by such a modification. This indicates that metalinguistic demand can’t account for the kind of differences we observed, contrary to our proposed guided vs. spontaneous dichotomy.

This could also help understand the somewhat small effect sizes observed in pragmatic research, compared to the challenges that autistic individuals and their relatives report in clinical practice. Developing experimental paradigms that are sensitive to spontaneous pragmatic reasoning could thus help in detecting the difficulties that autistic individuals can

experience in real-life interactions which, compared to lab or clinic-controlled environment, additionally demand to keep track of others’ minds in much noisier situations. Interestingly, such a distinction between guided and spontaneous pragmatics processes echoes a proposal in the theory-of-mind literature, in which autistic individuals were initially thought to have a lack of competence (Baron-Cohen, 1997) but which was later contested, resulting in a subtler proposition of some present competences but a lack of *spontaneous* use of those (Senju, 2012).

Of course, this hypothesis remains an *a posteriori* explanation, which should now be specifically tested, for instance by comparing how autistic individuals perform in a guided vs spontaneous measure of a single pragmatic phenomena, using carefully controlled materials and methodology which also controls for language demands. More generally, our study presents some general limitations that should be addressed in future research. It is possible that the patterns we observed evolve with age, but our sample size limits the conclusion we can draw on age-related differences. Increasing sample size might also reveal subtler effects, such as non-linear developmental patterns, or slighter effect on response times, which measure prompt important noise. Moreover, we carefully selected autistic participants to control for a series of potential confounds, but in real-life autism is most frequently associated with other psychiatric or neurodevelopmental conditions (Levy et al., 2010); the generalizability of our results to other subgroups of autistic individuals thus still needs to be assessed.

## **Conclusion**

In this study, we tested a group of autistic children with unimpaired language and non-verbal skills on two different pragmatic tasks that involved metaphors, which are

classically considered as difficult for autistic individuals, and scalar implicatures, with which autistic participants have mostly been found to perform typically. Yet, compared to typically developing children, our clinical sample showed a reversed contrast, with poorer scalar implicatures comprehension but typical metaphor comprehension. Accuracy and response times pattern were consistent with that view. Complementary to previously introduced dichotomies, we suggest that this reversed contrast could be explained by an opposition of *guided vs spontaneous pragmatic processes*. Paralleling theory of mind hypothesis, autistic individuals could show typical pragmatic abilities, as measured with guiding paradigms, but might struggle to recruit those skills spontaneously. This would result in poorer performances in tasks which do not guide pragmatic reasoning, and difficulties in everyday life interactions, which additionally demand to keep track of others' minds in a much noisier environment.

## References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). American Psychiatric Publishing.
- 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, 996. <https://doi.org/10.3389/fpsyg.2017.00996>
- Andrés-Roqueta, C., & Katsos, N. (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](https://doi.org/10.1044/2020_JSLHR-19-00263)
- Asperger, H. (1944). Die „Autistischen psychopathen“ im Kindesalter (U. Frith, Trans.). *Archiv Für Psychiatrie Und Nervenkrankheiten*, 117(1), 37–92.
- Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/doi:10.18637/jss.v067.i01>.
- Botha, M., Hanlon, J., & Williams, G. L. (2021). Does Language Matter? Identity-First Versus Person-First Language Use in Autism Research: A Response to Vivanti. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-020-04858-w>
- Chahboun, S., Vulchanov, V., Saldaña, D., Eshuis, H., & Vulchanova, M. (2016). Can You Play with Fire and Not Hurt Yourself? A Comparative Study in Figurative Language Comprehension between Individuals with and without Autism Spectrum Disorder. *PloS One*, 11(12), e0168571. <https://doi.org/10.1371/journal.pone.0168571>
- Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar Inferences in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 40(9), 1104–1117. <https://doi.org/10.1007/s10803-010-0960-8>
- Chouinard, B., & Cummine, J. (2016). All the world's a stage: Evaluation of two stages of metaphor comprehension in people with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 23, 107–121. <https://doi.org/10.1016/j.rasd.2015.12.008>
- Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) Family Affluence Scale. *Social Science & Medicine*, 66(6), 1429–1436. <https://doi.org/10.1016/j.socscimed.2007.11.024>
- Deckert, M., Schmoeger, M., Schaunig-Busch, I., & Willinger, U. (2019). Metaphor processing in middle childhood and at the transition to early adolescence: The role of chronological age,

- mental age, and verbal intelligence. *Journal of Child Language*, 46(2), 334–367. <https://doi.org/10.1017/S0305000918000491>
- 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>
- Gold, R., & Faust, M. (2010). Right hemisphere dysfunction and metaphor comprehension in young adults with Asperger syndrome. *Journal of Autism and Developmental Disorders*, 40(7), 800–811. <https://doi.org/10.1007/s10803-009-0930-1>
- Gold, R., Faust, M., & Goldstein, A. (2010). Semantic integration during metaphor comprehension in Asperger syndrome. *Brain and Language*, 113(3), 124–134. <https://doi.org/10.1016/j.bandl.2010.03.002>
- Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48(2), 101–119.
- Happé, F., Cook, J. L., & Bird, G. (2017). The Structure of Social Cognition: In(ter)dependence of Sociocognitive Processes. *Annual Review of Psychology*, 68(1), 243–267. <https://doi.org/10.1146/annurev-psych-010416-044046>
- Hermann, I., Haser, V., van Elst, L. T., Ebert, D., Müller-Feldmeth, D., Riedel, A., & Konieczny, L. (2013). Automatic metaphor processing in adults with Asperger syndrome: A metaphor interference effect task. *European Archives of Psychiatry and Clinical Neuroscience*, 263(S2), 177–187. <https://doi.org/10.1007/s00406-013-0453-9>
- 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>
- Horn, L. R. (1989). *A natural history of negation* (University of Chicago Press).
- Kalandadze, T., Bambini, V., & Næss, K.-A. B. (2019). A systematic review and meta-analysis of studies on metaphor comprehension in individuals with autism spectrum disorder: Do task properties matter? *Applied Psycholinguistics*, 40(6), 1421–1454. <https://doi.org/10.1017/S0142716419000328>
- Kalandadze, T., Norbury, C., Naerland, T., & Naess, K.-A. B. (2018). Figurative Language Comprehension in Individuals with Autism Spectrum Disorder: A Meta-Analytic Review. *Autism: The International Journal of Research and Practice*, 22(2), 99–117. <https://doi.org/10.1177/1362361316668652>
- Kanner, L. (1946). Irrelevant and metaphorical language in early infantile autism. *The American Journal of Psychiatry*, 103(2), 242–246. <https://doi.org/10.1176/ajp.103.2.242>
- Katsos, N. (2014). Scalar implicature. *Pragmatic Development in First Language Acquisition*, 10, 183–197.
- Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, 120(1), 67–81. <https://doi.org/10.1016/j.cognition.2011.02.015>
- Katsos, N., Cummins, C., Ezeizabarrena, M.-J., Gavarró, A., Kuváč Kraljević, J., Hrzica, G., Grohmann, K. K., Skordi, A., Jensen de López, K., Sundahl, L., van Hout, A., Hollebrandse, B., Overweg, J., Faber, M., van Koert, M., Smith, N., Vija, M., Zupping, S., Kunnari, S., ... Noveck, I. (2016). Cross-linguistic patterns in the acquisition of quantifiers. *Proceedings of the National Academy of Sciences*, 113(33), 9244–9249. <https://doi.org/10.1073/pnas.1601341113>
- Khomsí, A., Khomsí, F., & Pasquet, F. (2007). *BILO - Bilans Informatisés de Langage Oral*. Pearson.
- Kissine, M., & De Brabanter, P. (2023). Pragmatic responses to under-informative some-statements are not scalar implicatures. *Cognition*, 237, 105463. <https://doi.org/10.1016/j.cognition.2023.105463>
- Lenth, R. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (R package version 1.8.2). <https://CRAN.R-project.org/package=emmeans>

- Levy, S. E., Giarelli, E., Lee, L.-C., Schieve, L. A., Kirby, R. S., Cunniff, C., Nicholas, J., Reaven, J., & Rice, C. E. (2010). Autism Spectrum Disorder and Co-occurring Developmental, Psychiatric, and Medical Conditions Among Children in Multiple Populations of the United States. *Journal of Developmental & Behavioral Pediatrics, 31*(4), 267. <https://doi.org/10.1097/DBP.0b013e3181d5d03b>
- Marocchini, E. (2023). Impairment or difference? The case of Theory of Mind abilities and pragmatic competence in the Autism Spectrum. *Applied Psycholinguistics, 1–19*. <https://doi.org/10.1017/S0142716423000024>
- Mazzaggio, G., Foppolo, F., Job, R., & Surian, L. (2021). Ad-hoc and scalar implicatures in children with autism spectrum disorder. *Journal of Communication Disorders, 90*, 106089. <https://doi.org/10.1016/j.jcomdis.2021.106089>
- Morsanyi, K., Hamilton, J., Stamenković, D., & Holyoak, K. J. (2022). Linking metaphor comprehension with analogical reasoning: Evidence from typical development and autism spectrum disorder. *British Journal of Psychology, 113*(2), 479–495. <https://doi.org/10.1111/bjop.12542>
- Morsanyi, K., Stamenković, D., & Holyoak, K. J. (2020). Metaphor processing in autism: A systematic review and meta-analysis. *Developmental Review, 57*, 100925. <https://doi.org/10.1016/j.dr.2020.100925>
- Müller, E., Schuler, A., & Yates, G. B. (2008). Social challenges and supports from the perspective of individuals with Asperger syndrome and other autism spectrum disabilities. *Autism, 12*(2), 173–190. <https://doi.org/10.1177/1362361307086664>
- 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>
- Pastor-Cerezuela, G., Tordera Yllescas, J. C., González-Sala, F., Montagut-Asunción, M., & Fernández-Andrés, M.-I. (2018). Comprehension of Generalized Conversational Implicatures by Children With and Without Autism Spectrum Disorder. *Frontiers in Psychology, 9*, 272. <https://doi.org/10.3389/fpsyg.2018.00272>
- Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J.-P., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders, 39*(4), 607–618. <https://doi.org/10.1007/s10803-008-0661-8>
- Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics, 25*, 51–79. <https://doi.org/10.1075/bjl.25.04pou>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526. <https://doi.org/10.1017/S0140525X00076512>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rundblad, G., & Annaz, D. (2010). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism: The International Journal of Research and Practice, 14*(1), 29–46. <https://doi.org/10.1177/1362361309340667>
- Schaeken, W., Van Haeren, M., & Bambini, V. (2018). The Understanding of Scalar Implicatures in Children With Autism Spectrum Disorder: Dichotomized Responses to Violations of Informativeness. *Frontiers in Psychology, 9*, 1266. <https://doi.org/10.3389/fpsyg.2018.01266>
- Semino, E. (2014). Pragmatic failure, mind style and characterisation in fiction about autism. *Language and Literature (Harlow, England), 23*(2), 141–158. <https://doi.org/10.1177/0963947014526312>
- Senju, A. (2012). Spontaneous Theory of Mind and Its Absence in Autism Spectrum Disorders. *The Neuroscientist, 18*(2), 108–113. <https://doi.org/10.1177/1073858410397208>
- Su, Y., & Su, L.-Y. (2015). Interpretation of Logical Words in Mandarin-Speaking Children with Autism Spectrum Disorders: Uncovering

Knowledge of Semantics and Pragmatics. *Journal of Autism and Developmental Disorders*, 45(7), 1938–1950. <https://doi.org/10.1007/s10803-014-2350-0>

Surian, L., Baron-Cohen, S., & van der Lely, H. (1996). Are Children with Autism Deaf to Gricean Maxims? *Cognitive Neuropsychiatry*, 1(1), 55–72. <https://doi.org/10.1080/135468096396703>

Thomas, M. S. C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using Developmental Trajectories to Understand Developmental Disorders. *Journal of Speech, Language, and Hearing Research*, 52(2), 336–358. [https://doi.org/10.1044/1092-4388\(2009/07-0144\)](https://doi.org/10.1044/1092-4388(2009/07-0144))

Van Herwegen, J., & Rundblad, G. (2018). A Cross-Sectional and Longitudinal Study of Novel Metaphor and Metonymy Comprehension in Children, Adolescents, and Adults With Autism Spectrum Disorder. *Frontiers in Psychology*, 9, 945. <https://doi.org/10.3389/fpsyg.2018.00945>

Vivanti, G. (2020). Ask the Editor: What is the Most Appropriate Way to Talk About Individuals with a Diagnosis of Autism? *Journal of Autism and Developmental Disorders*, 50(2), 691–693. <https://doi.org/10.1007/s10803-019-04280-x>

Wilson, A. C. (2022). Development and validation of the conversation questionnaire: A psychometric measure of communication challenges generated from the self-reports of autistic people. *Autism & Developmental Language Impairments*, 7, 239694152211232. <https://doi.org/10.1177/23969415221123286>

Wilson, A. C., & Bishop, D. V. M. (2021). “Second guessing yourself all the time about what they really mean...”: Cognitive differences between autistic and non-autistic adults in understanding implied meaning. *Autism Research*, 14(1), 93–101. <https://doi.org/10.1002/aur.2345>

Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology*, 12, 289–297. <https://doi.org/10.1037/0012-1649.12.4.289>

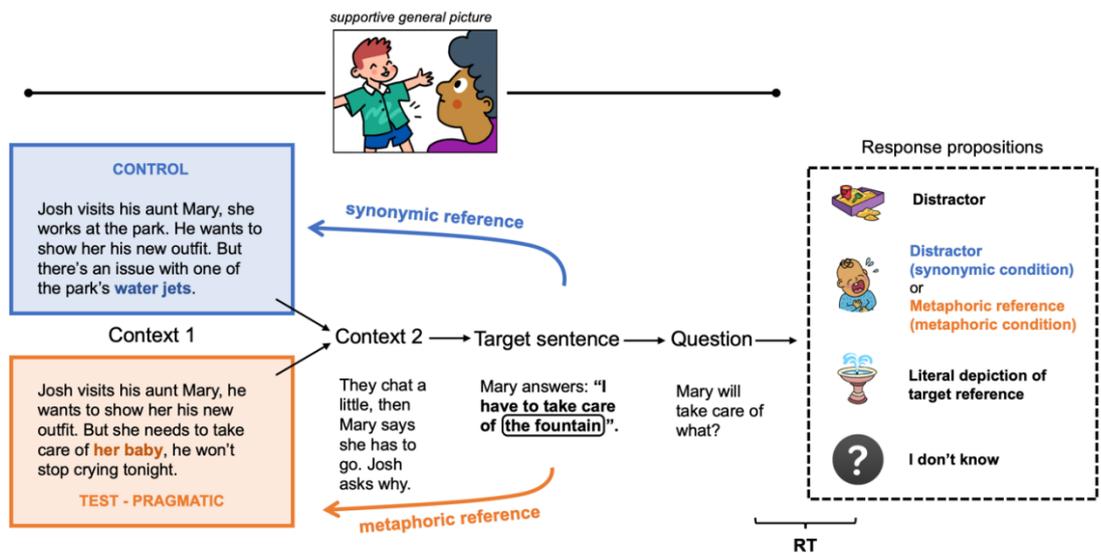
## Supplementary materials

**Table S1: Clinical group detailed description**

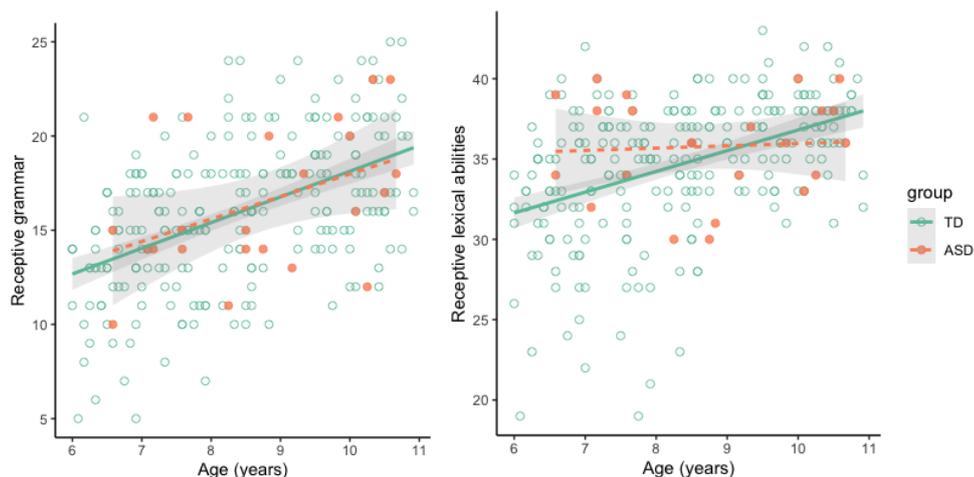
<b>ID</b>	<b>Age (years; months)</b>	<b>Class</b>	<b>Sex</b>	<b>SES</b>	<b>Autism Assessment</b>	<b>Intelligence assessment*</b>
<b>1</b>	6;7	1 <sup>st</sup> grade	M	80	positive ADOS-2; positive SCQ	FSIQ = 105 (WPPSI-IV)
<b>2</b>	6;7	1 <sup>st</sup> grade	M	17	positive ADOS-2; negative SCQ	<i>NA</i>
<b>3</b>	7;1	1 <sup>st</sup> grade	M	50	positive SCQ; positive ADI-R	VC = 100; VS = 102; FR = 109 (WISC-V)
<b>4</b>	7;2	1 <sup>st</sup> grade	M	20	positive SCQ	FSIQ = 131 (WPPSI-IV)
<b>5</b>	7;2	1 <sup>st</sup> grade	M	65	positive SCQ	FSIQ = 138 (WISC-V)
<b>6</b>	7;7	2 <sup>nd</sup> grade	M	65	positive ADOS-2; positive SCQ	VC = 98; VS = 102; FR = 112 (WISC-V)
<b>7</b>	7;7	2 <sup>nd</sup> grade	F	43	positive SCQ	<i>NA</i>
<b>8</b>	7;8	3 <sup>rd</sup> grade	M	35	positive ADOS-2; positive SCQ	VC = 114; VS = 117; FR = 135 (WPPSI-VI)
<b>9</b>	8;3	2 <sup>nd</sup> grade adapted	F	24	positive ADOS-2; positive SCQ	VC = 92; VS = 89; FR = 97 (WISC-V)
<b>10</b>	8;6	3 <sup>rd</sup> grade	M	49	positive ADOS-2; positive SCQ	VC = 95; VS = 105; FR = 118 (WISC-V)
<b>11</b>	8;6	3 <sup>rd</sup> grade	M	54	positive ADOS-2	FSIQ = 105 (WISC-V)
<b>12</b>	8;9	3 <sup>rd</sup> grade	M	84	positive ADOS-2; positive SCQ	FSIQ = 79 (WPPSI-IV)
<b>13</b>	8;10	4 <sup>th</sup> grade	M	60	positive ADOS-2; negative SCQ	FSIQ = 124 (WPPSI-IV)
<b>14</b>	9;2	3 <sup>rd</sup> grade adapted	M	54	positive ADOS-2	FSIQ = 88 (WNV)
<b>15</b>	9;4	4 <sup>th</sup> grade	M	58	positive ADOS-2; positive SCQ	VC = 98; VS = 102; FR = 112 (WISC-V)
<b>16</b>	9;10	4 <sup>th</sup> grade	M	50	positive ADOS-2; positive SCQ	FSIQ = 92 (WISC-V)
<b>17</b>	10;0	4 <sup>th</sup> grade	F	62	positive ADOS-2; positive SCQ	FSIQ = 100 (WISC-V)
<b>18</b>	10;1	4 <sup>th</sup> grade	M	61	positive ADOS-2; positive SCQ	VC = 118; VS = 138; FR = 118 (WISC-V)
<b>19</b>	10;3	4 <sup>th</sup> grade adapted	M	72	positive SCQ	<i>NA</i>
<b>20</b>	10;4	5 <sup>th</sup> grade	M	47	positive ADOS-2; negative SCQ	FSIQ = 112 (WISC-V)
<b>21</b>	10;6	5 <sup>th</sup> grade	M	56	positive ADOS-2; negative SCQ	FSIQ = 115 (WPPSI-IV)
<b>22</b>	10;7	5 <sup>th</sup> grade	M	64	positive SCQ	FSIQ = 127 (WISC-V)
<b>23</b>	10;8	5 <sup>th</sup> grade	M	54	positive ADOS-2	FSIQ = 121 (WISC-V)

\* *FSIQ is reported when available. If not, other available summary values are reported. The tests used to provide these measures are detailed.*

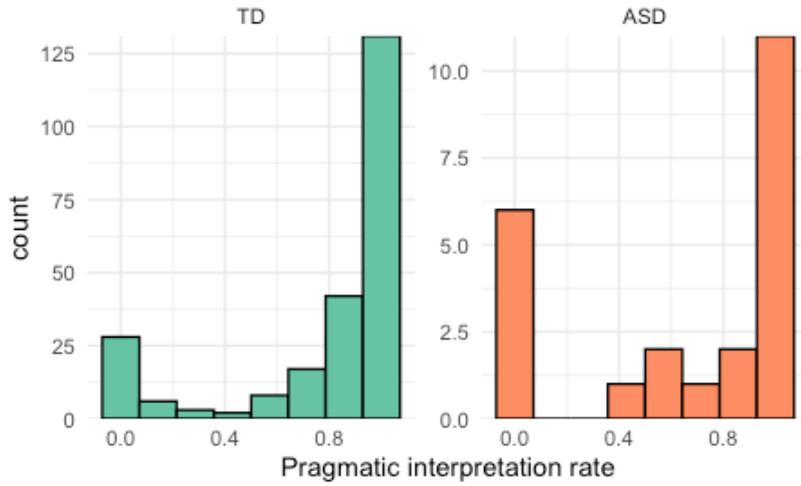
*Abbreviations used: SCQ = Social Communication Questionnaire; ADOS-2 = Autism Diagnostic Observation Scale 2<sup>nd</sup> Edition; FSIQ = Full Scale Intelligence Quotient; VC = Verbal Comprehension; VS = Visuo-Spatial; FR = Fluid Reasoning; WPPSI-IV = Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition; WISC-V = Wechsler Intelligence Scale for Children, Fifth Edition; WNV= Wechsler Nonverbal Scale of Ability*



**Figure S1:** représentation de la structure d'un item de la métaphore tâche (traduit de français)



**Figure S2:** Scatter plots of all participants' performances and linear regressions of receptive grammar (left) and receptive lexicon (right) against age and group, revealing no significant group main effect or interaction



**Figure S3:** Histogram of individual pragmatic interpretation rates in each group, revealing bimodal distributions in both cases

**Table S2:** Output from the main model explaining response times in the metaphor task

<b>LOG(RT)</b>			
<b>PREDICTORS</b>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
<b>(INTERCEPT)</b>	7.99	7.93 - 8.05	<b>&lt;0.001</b>
<b>GROUP</b>	0.21	0.11 - 0.32	<b>&lt;0.001</b>
<b>CONDITION</b>	-0.16	-0.27 - -0.04	<b>0.007</b>
<b>GROUP * CONDITION</b>	0.03	-0.10 - 0.15	0.643
<b>RANDOM EFFECTS</b>			
$\Sigma^2$	0.15		
$T_{00}$ PARTICIPANT	0.04		
$T_{00}$ ITEMID	0.01		
<b>ICC</b>	0.25		
<b>N</b> ITEMID	12		
<b>N</b> PARTICIPANT	260		
<b>OBSERVATIONS</b>	2125		
<b>MARGINAL R<sup>2</sup> / CONDITIONAL R<sup>2</sup></b>	0.045 / 0.282		

## Annexe E • Table des figures

<b>Figure 1-1 :</b> <i>Modèle du code (Shanon, 1948)</i>	18
<b>Figure 1-2 :</b> <i>Représentation de la pragmatique au sens fort, comme base cognitive du langage (reproduction et traduction de Scott-Phillips, 2017)</i>	20
<b>Figure 1-3 :</b> <i>Exemple de configuration de points complexe (A) ou simple (B) à mémoriser pendant la tâche d'implicature scalaires (illustration créée d'après De Neys &amp; Schaeken, 2007)</i>	27
<b>Figure 1-4 :</b> <i>Exemple d'item utilisé par Huang et Snedecker (2009a), dans lequel une instruction critique serait « Point to the girl who has some of the socks », comparé à « Point to the girl who has all of the socker balls »</i>	29
<b>Figure 1-5 :</b> <i>Potentiels évoqués par le quantifieur 'quelques' dans l'étude de Zhao et al. (2021), pour les répondeurs pragmatiques dans la condition 'focus' (à gauche) où sont visibles des composants P200, N400 et la négativité soutenue (SN) plus marquées en réponses aux quantifieur sous-informatifs (rouge) qu'informatifs (bleu), et les répondeurs sémantiques pour la condition 'non-focus' (à droite), où de telles différences sont absentes (figure extraite de la publication originale)</i>	31
<b>Figure 1-6 :</b> <i>Temps de lecture des participants de l'étude d'Ortony et al. (1978) pour les interprétations des expressions critiques, selon que le contexte induise une interprétation métaphorique ou littérale, et que le contexte soit pauvre ou riche (figure extraite de la publication originale)</i>	33
<b>Figure 1-7 :</b> <i>Potentiels évoqués par des stimuli métaphoriques vs littéraux dans un contexte minimal dans l'étude de Bambini et al. (2016, expérience 1)</i>	35
<b>Figure 1-8 :</b> <i>Effet facilitateur de la métaphore sur le traitement de mots liés aux propriété pertinentes (distinctive properties) ou non-pertinentes (superordinates) par rapport aux mots non-liés, à différents délais après la présentation de la métaphore (ISI) (figure extraite de Rubio-Fernandez, 2007)</i>	37
<b>Figure 1-9 :</b> <i>Illustration des processus et comportements multiples qui sont regroupés sous le concept de théorie de l'esprit et que Schaafsma et al. (2015) appellent à déconstruire, puis à reconstruire (extrait de la publication originale)</i>	40
<b>Figure 1-10 :</b> <i>Modèle de l'interprétation du langage d'A.C. Wilson et Bishop (2019) reprenant la théorie de la pertinence (traduction personnelle)</i>	42
<b>Figure 1-11 :</b> <i>Figure extraite de Spotorno et al. (2012) et représentant la suractivation cérébrale associée aux énoncés ironiques par rapport aux énoncés littéraux dans 4 régions cérébrales typiquement associées à la théorie de l'esprit (cortex préfrontal médian = MPFC, jonctions temporo-pariétales gauche et droite = lTPJ et rTPJ et précunéus = PC).</i>	44
<b>Figure 2-1 :</b> <i>Stades d'interprétation des métaphores selon Winner et al. (1976) avec des exemples d'interprétation de la métaphore "L'odeur de son parfum était un rayon de soleil" (à droite), et la fréquence des réponses de type métaphorique complet ou primitif, en modalité d'explication verbale libre, en fonction de l'âge (valeurs extraites de la publication originale)</i>	55
<b>Figure 2-2 :</b> <i>Taux de réponses correctes (gauche, expérience 1) et temps de réponse (droite, expérience 2) en fonction de l'âge à la tâche de références métaphoriques proposée par Noveck et al. (2001, données extraites de la publication originale ; la moyenen d'âge du groupe d'adultes a été utilisée pour l'expérience 2)</i>	58

<b>Figure 2-3 :</b> Trajectoires développementales des performances des enfants dans l'étude de Derkert et al. (2019) pour les scores d'identification (a), d'explication (b) et de préférence (c), mettant en évidence les patterns non-linéaires (en pointillés) expliquant mieux les données que des patterns linéaires (traits pleins), avec les intervalles de confiance à 95% pour les âges pivots encadrés (figure extraite de la publication originale)	60
<b>Figure 2-4 :</b> Illustration de l'item critique du paradigme basé sur l'action proposé par Pouscoulous et al. (2007, expérience 2, gauche) et proportion de chaque groupe répondant de façon pragmatique (droite) (données extraites de la publication originale)	65
<b>Figure 2-5 :</b> Représentation du paradigme utilisé par Katsos et Bishop (2011, en haut) et des résultats obtenus par les groupes aux items critiques en fonction de la modalité de réponse (binaire ou ternaire, en bas ; données extraites de la publication originale pour la condition scalaire uniquement)	68
<b>Figure 3-1 :</b> Illustration des perceptions induites par différents choix terminologiques pour évoquer les personnes ayant un diagnostic d'autisme, traduite d'un tweet d'@Autistic_Cherry	72
<b>Figure 3-2 :</b> Différents niveaux d'explication du neurodéveloppement, avec les trois niveaux classiquement considérés pour caractériser l'autisme (en vert), et le niveau sociétal proposé par les modèles sociologiques (en bleu) (figure adaptée de Fletcher-Watson & Happé, 2019, reproduction autorisée par Taylor and Francis Group, LLC, une filiale de Informa plc)	73
<b>Figure 3-3 :</b> Données d'eye-tracking d'adultes neurotypiques (NT) et autistes (AS) dans une tâche de fausses croyances implicites, représentant le rapport entre le temps de fixation de la cible et celui du distracteur (A) et la proportion de premiers regards vers la cible plutôt que le distracteur (B) (lignes pointillées = niveau de chance ; figure extraite de Senju, 2012)	76
<b>Figure 3-4 :</b> Performances en compréhension de métaphores des enfants autistes et neurotypiques (TD) en fonction de l'âge chronologique (a), de l'âge mental verbal (b) et de l'âge mental non-verbal (c) dans l'étude de Rundblad et Annaz (2010 ; figures extraites de la publication originale)	83
<b>Figure 3-5 :</b> Représentation des étapes cognitives de la compréhension d'une métaphore d'après le modèle de Glucksberg et al. (1982), dont l'interférence métaphorique témoignerait (figure traduite de Chouinard et al., 2017)	85
<b>Figure 3-6 :</b> Représentation visuelle des analyses de connectivité extraite de Chouinard et al. (2017), représentant les liens significatif (traits pleins) entre les différentes régions d'intérêt considérées, incluant des régions corticales (1 à 7) et sous-corticales (A à F). 1 = gyrus temporal médian, 2 = gyrus temporal supérieur, 3 = cortex préfrontal dorsolatéral, 4 = cortex cingulaire antérieur, 5 = gyrus frontal inférieur, 6 = cortex pariétal postérieur droit, 7 = insula ; A = noyau subthalamique, B = putamen, C = thalamus, D = globus pallidus interne, E = globus pallidus externe, F = noyau caudé	87
<b>Figure 3-7 :</b> Temps de réponse observé par Chevallier et al. (2010) chez les participants autistes (ASD, en jaune) ou non (TD, en blanc) dans la condition critique (encadrée en vert) ou les conditions contrôles (couleurs et encadré ajoutés à la figure originale)	89
<b>Figure 3-8 :</b> Illustration du paradigme de connaissance partielle d'Hochstein et al. (2017) avec les réponses attendues aux items critiques (en couleur) en fonction de la condition	90
<b>Figure 3-9 :</b> Exemple d'item critique avec une implicature scalaire, extrait de Mazzaggio et al. (2021) : la réponse attendue est l'image en bas à droite, tandis que l'image en bas à gauche est une réponse possible en cas d'interprétation logique (support visuel extrait de la publication originale, consigne orale traduite depuis l'italien)	92
<b>Figure 3-10 :</b> Illustration du processus d'interprétation du langage d'après la théorie de la pertinence (Sperber, 2005; Sperber & Wilson, 1996), tel que décrit par Mazzarella et Noveck (2021)	96
<b>Figure 4-1 :</b> Initial display of an item of the task	109
<b>Figure 4-2 :</b> Items distribution in the scalar task	110

<b>Figure 4-3:</b> Distribution of the rates of pragmatic interpretations to critical items of each child (left) and their evolution with age (right: children's individual pragmatic interpretation rates as a function of age, described with a LOESS regression line (plain yellow, SE in grey) and a linear regression line (dashed black)	113
<b>Figure 4-4:</b> Group responses to underinformative items for each item position; note that the first and the sixth item are identical	114
<b>Figure 4-5:</b> Predicted response times for each answer-type at different ages (partial effects plot, with confidence intervals in shades)	115
<b>Figure 5-1:</b> Illustration of a vignette pair structure (translated from French to English) depicting what a participant sees (top) and hears (bottom)	130
<b>Figure 5-2:</b> Proportion rates of each answer for the different conditions (based on raw data), across the different age groups, with accuracy for each age group in each condition	140
<b>Figure 5-3:</b> Predicted accuracy in the different conditions, as a function of TOM (x-axis, above) and grammar (x-axis, below), at 3 different ages (6-, 8- and 10-year-olds) (predictions were backtransformed from scaled variables)	142
<b>Figure 5-4:</b> Predicted response times (in ms, on a logarithmic scale) as a function of age and condition, and partial residuals (residuals below 1000 ms or above 8000 ms [i.e., 1% of the data points] were not displayed and age was backtransformed from scaled values to improve readability).	143
<b>Figure 5-5:</b> Predicted response times (in ms, log-transformed) across the two reference conditions, as a function of TOM (x-axis, top half) and grammar (x-axis, bottom half), at 3 different age points (6, 8 and 10 years old) (predictions were backtransformed from scaled variables).	144
<b>Figure 5-6:</b> Predicted partial effects of TOM and language on accuracy (left) and response times for correct responses (right) in the metaphoric reference condition, as a function of age (note that the y axis in (b) reverses polarity to indicate increased speed).	145
<b>Figure 6-1 :</b> performances individuelles et régressions linéaires des participants aux épreuves de compréhension grammaticale (gauche) et lexicale (droite) en fonction de l'âge et du groupe ; IC à 95% représentés par les zones grisées	159
<b>Figure 6-2 :</b> Probabilité des réponses pragmatiques prédites par le modèle statistique en fonction de l'âge et du groupe	160
<b>Figure 6-3 :</b> Effets partiels du groupe et du type de réponse sur les temps de réponse (après transformation logarithmique) d'après le modèle statistique ; les barres d'erreurs représentent les intervalles de confiance à 95%	161
<b>Figure 6-4 :</b> Probabilité de réponse correcte prédite par le modèle généralisé en fonction du groupe, de l'âge et de la condition	162
<b>Figure 6-5 :</b> Temps de réponse (après transformation logarithmique) prédits par le modèle en fonction du groupe et de la condition; les barres d'erreurs représentent les intervalles de confiance à 95%	162
<b>Figure 7-1 :</b> Principaux résultats de l'étude 1, représentant l'évolution avec l'âge des taux de réponses pragmatiques individuels des enfants tout-venant (gauche) et de leurs temps de réponse en fonction du type de réponse (droite)	171
<b>Figure 7-2 :</b> Effets partiels des prédicteurs cognitifs (langage et théorie de l'esprit) sur les performances (à gauche) et les temps de réponse (à droite) des enfants dans la condition test de la tâche de métaphores, en fonction de l'âge (étude 2, expérience 2)	173
<b>Figure 7-3 :</b> Trajectoire développementales des réponses des groupes neurotypique et autistique dans les conditions tests des tâches d'implicatures scalaires (à gauche) et de métaphores (à droite), issue de l'étude 3	174
<b>Figure 7-4 :</b> Positionnement des tâches de cette thèse et d'autres tâches de la littérature sur un axe opposant des raisonnements pragmatiques guidés vs spontanés, et impactant les performances de participants autistes (TSA) relativement à des contrôles neurotypiques (NT)	180

<b>Figure 7-5 :</b> Rapport entre les coûts et les bénéfices estimés d'une information, suite par exemple à la présentation du quantifieur 'quelques'. À gauche, le rapport est équilibré et pousse à inférer les intentions de communication déterminant l'énoncé (ici, en comprenant l'implicature scalaire). À droite, le rapport est déséquilibré et ne guide pas l'inférence, ce qui peut résulter d'une surévaluation des coûts et/ou d'une sous-estimation des bénéfices	183
<b>Figure 7-6 :</b> Représentation du principe de base des modèles bayésiens, extraite et traduite de Haker et al. (2016), représentant la comparaison des prédictions d'un modèle interne à l'input sensoriel issu de l'environnement et aboutissant à une perception et à des erreurs de prédiction que le système vise à minimiser.	184
<b>Figure 7-7 :</b> Distribution pour chacun des groupes de l'étude 3 du score composite individuel indexant l'effort cognitif associé aux implicatures scalaires, et seuil optimal (ligne pointillée) discriminant les deux groupes, avec les valeurs associées de sensibilité et de spécificité	187
<b>Figure B-1 :</b> PST stories example, involving false beliefs to assess theory of mind (above), or mechanical causalities and social scripts as control (below)	227
<b>Figure B-2 :</b> Task set up for a practice item as it appears on the tablets of the participants	229
<b>Figure B-3 :</b> model's predicted accuracy as a function of age and item-type (shaded areas represent 95% confidence intervals)	233
<b>Figure B-4 :</b> GSA standardized index' distribution for the autistic and the control group	238
<b>Figure B-5 :</b> Joint distributions of the TOM-Accuracy and the TOM-RT indices for the control group (left panel) and the autistic group (right panel).	238
<b>Figure B-6 :</b> Distribution across groups (left panel) and ROC curve (right panel) of the TOM-Effort index	239
<b>Figure D-1 :</b> predicted probability of pragmatic responses in the scalar task based on age and group	268
<b>Figure D-2 :</b> Response-type and group partial effects and interaction on response times in the scalar task; error bars represent 95% CI.	269
<b>Figure D-3 :</b> predicted accuracy in the metaphor task for both groups based on age and condition	270
<b>Figure D-4 :</b> predicted response times in the metaphor task for both groups in both conditions	271

*NB : Les figures des matériels supplémentaires des articles (chapitres 4, 5, annexes B, D) n'ont pas été référencées dans cette table.*

## **Annexe F • Serment du docteur**

*En application de l'article L612-7 du code de l'éducation, les établissements d'enseignement supérieur et de recherche mettent en œuvre, dans le cadre des soutenances de leurs doctorantes et doctorants, une prestation de serment d'intégrité scientifique. Cette disposition est inscrite dans l'article 19bis de l'arrêté de la formation doctorale (évolution introduite par l'arrêté du 26 août 2022 modifiant l'arrêté du 25 mai 2016).*

En présence de mes pairs.

Parvenu à l'issue de mon doctorat en sciences cognitives, et ayant ainsi pratiqué, dans ma quête du savoir, l'exercice d'une recherche scientifique exigeante, en cultivant la rigueur intellectuelle, la réflexivité éthique et dans le respect des principes de l'intégrité scientifique, je m'engage, pour ce qui dépendra de moi, dans la suite de ma carrière professionnelle quel qu'en soit le secteur ou le domaine d'activité, à maintenir une conduite intègre dans mon rapport au savoir, mes méthodes et mes résultats.

