Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/plcp21

Scalar inference is supported by Theory of Mind networks in adults and children

Alyssa F. Kampa, Anna Papafragou & Kaja K. Jasińska

To cite this article: Alyssa F. Kampa, Anna Papafragou & Kaja K. Jasińska (10 Apr 2025): Scalar inference is supported by Theory of Mind networks in adults and children, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2025.2489602

To link to this article: https://doi.org/10.1080/23273798.2025.2489602

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2025.2489602

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Scalar inference is supported by Theory of Mind networks in adults and children

Alyssa F. Kampa[a], Anna Papafragou[b] and Kaja K. Jasińska[c,d]

[a]Department of Linguistics, University of Delaware, Newark, DE, USA; [b]Department of Linguistics, University of Pennsylvania, Philadelphia, PA, USA; [c]Applied Psychology and Human Development, University of Toronto, Toronto, Canada; [d]Haskins Laboratories, New Haven, CT, USA

ABSTRACT

Scalar implicatures, a type of pragmatic inference that relies on the evaluation of alternatives on a logical scale, have been extensively studied in the developmental literature, yet their developmental timeline remains hazy. Furthermore, debates continue over the contributions that potential supporting factors, such as Theory of Mind, executive functions, and language, make to the scalar implicature derivation process in adulthood and during development. We present a novel approach to address these issues: we tested 4- and 5-year-old children (majority white with college-educated mothers and slightly higher than average SES) and adults (undergraduate students) on scalar implicature and theory of mind tasks using spatial neuroimaging techniques (fNIRS). We find evidence that neural networks associated with Theory of Mind, executive functions, and language were active during scalar inference in adults and some preschool-aged children. Moreover, we find that children who pass the scalar inference task show activation of neural networks associated with Theory of Mind and language processing (specifically including the dmPFC and LIFG) during scalar inference and right temporoparietal junction activity during a Theory of Mind task, while children who do not pass the task do not show activation of these regions. This study provides the first exploration into neural correlates of scalar inference in 4- and 5-year-olds using spatial neuroimaging techniques.

ARTICLE HISTORY
Received 5 June 2024
Accepted 24 March 2025

KEYWORDS
Pragmatics; scalar implicature; neuroimaging; Theory of Mind; language acquisition

Having a conversation with someone often requires the ability to go beyond what is said to interpret what is meant. For example, if at a garage sale someone says, "Some toys on this table are free," an adult knows to interpret this as "Some but not all toys on this table are free." These intended meanings are rarely, if ever, spoken out loud, and adults are never explicitly taught to connect these meanings with a specific utterance, yet even adults with wildly different backgrounds interpret them successfully and consistently. Given that children are unlikely ever to hear implied meanings spoken out loud, how do they develop the ability to make these inferences?

According to an influential model of communication (Grice, 1975), a listener can infer a speaker's intended meaning by utilising mutually available principles of communication that stem from expectations of rational behaviour. Rational communicators are expected to be cooperative to further the goals of a conversation (the *Cooperative Principle*), which requires adherence to certain communicative principles, or maxims, such as the *Maxim of Quantity*. The *Maxim of Quantity* states that a communicator should make their contribution as informative as required for the exchange (and no more and no less). For communication to be successful, both participants need to give as much information as is required of them cooperatively, and they assume that their conversational partner will do the same. Thus, when a speaker violates this communicative principle, such as by failing to give sufficient information, a listener might interpret their statement as having a non-literal intended meaning (to repair an otherwise unsuccessful contribution). In the previous example, given that the statement "All toys on this table are free" – if true – would give a greater amount of relevant information (information about all of the toys instead of just some), if a speaker instead states the less informative "Some toys on this table are free", the listener concludes that either 1) (the speaker knows that) "All of the toys on the table are free" is not true, or 2) the speaker does not know whether this more informative statement is true. If the listener can conclude that the speaker is knowledgeable about the status of the toys, e.g. she is in charge of the garage sale, then the listener will interpret "Some toys on the table are free" to mean "Not all toys on the table are free." This inference is referred to as a *scalar implicature*.

CONTACT Kaja K. Jasińska kaja.jasinska@utoronto.ca

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23273798.2025.2489602.

Much of the literature on scalar implicatures has focused on scalar implicatures derived from quantifiers. However, scalar implicatures can also be derived from other logically ordered scales, such as modals or numbers (Horn, 1989; Levinson, 2000). Scalar implicatures can even be derived from contextually-determined alternatives; these implicatures are called "particularized" or "ad-hoc" scalar implicatures (Hirschberg, 1985). If at the garage sale of the previous example, there is a teddy bear, a yo-yo, and a doll on the table, and the seller says, "The teddy bear and the yo-yo are free," the listener derives the inference that the doll is not free, following the same reasoning process about alternatives outlined above.

Experimental evidence for children's early competency with scalar inferences can sometimes appear conflicting, and debates continue over the exact developmental timeline for scalar implicature development and the factors that contribute to this development. Early studies report failure as late as 9-years (Noveck, 2001), while some recent studies report scalar implicature reasoning ability as early as 3.5-years (Kampa et al., 2024; Stiller et al., 2015). There is evidence that modifying the cognitive burden of a given task improves children's ability to successfully derive scalar implicatures, whether through simplifying task demands (Pouscoulous et al., 2007) or providing children with relevant alternatives (Skordos & Papafragou, 2016; Stiller et al., 2015).

We take the position, along with many studies, that the period between the ages of 3 and 5 often forms a critical juncture in the development of scalar implicatures (Barner et al., 2011; Chierchia et al., 2001; Foppolo et al., 2012; Guasti et al., 2005; Huang & Snedeker, 2009; Kampa & Papafragou, 2020, 2023; Noveck, 2001; Papafragou & Musolino, 2003; Stiller et al., 2015; among others). This period of development is also an important time for the development of other relevant cognitive skills and functions such as language (Gleason, 2005), executive functions (EF; Carlson, 2005; Garon et al., 2008; Zelazo & Müller, 2010), and Theory of Mind (ToM; Baron-Cohen et al., 1985; cf. Hollebrandse et al., 2014; Perner & Wimmer, 1985; Wellman et al., 2001; Wimmer & Perner, 1983), all of which may be recruited during the derivation of scalar implicatures in children and adults. Language development may be expected to interact with scalar implicature development not only because pragmatic inferences require the ability to produce and interpret language, but also because certain types of scalar implicatures, such as those arising from quantifiers, modals, and numerals, require knowledge of alternatives on a lexical scale. Indeed, there is evidence that linguistic factors, such as vocabulary and morpho-syntax, predict pragmatic ability (Foppolo et al., 2021; Rollins, 1999). EF skills also may be expected to play a role in scalar implicature derivation due to the heavy cognitive processing burden of the derivation process, including generating alternative statements the speaker could have said and rejecting them in light of the current discourse state, and there is some evidence that EF impact scalar implicature computation in adults (Antoniou et al., 2016; Apperly et al., 2008; De Neys & Schaeken, 2007; Marty & Chemla, 2013), although other studies have found conflicting evidence (Antoniou et al., 2014; Fairchild & Papafragou, 2021; Heyman & Schaeken, 2015).

There is considerably more controversy over the role of ToM in scalar implicature derivation. Under Gricean and several post-Gricean accounts, scalar implicature derivation is a complex, social, inferential process that requires the listener to incorporate information about the speaker's knowledge state and other characteristics (Grice, 1975, 1989; Sauerland, 2004; Sperber & Wilson, 1995/1986, 2002). However, under a different class of lexical and grammatical accounts, it has been proposed that scalar implicature derivation can be achieved through linguistic mechanisms alone without the need for epistemic reasoning (Chierchia, 2004, 2006; Chierchia et al., 2009, 2012; Fox, 2007; Levinson, 2000). The proposed differences in the nature of scalar implicature derivation between these two classes of theories necessarily give rise to different predictions about the requisite mechanisms and abilities involved. Evidence that ToM is recruited during scalar implicature derivation would not be directly accounted for by grammatical accounts. Conversely, evidence that ToM is not recruited during scalar implicature derivation in adults would pose serious problems for Gricean and neo-Gricean theories under which epistemic reasoning is thought to be a critical component of such a derivation. Relatedly, evidence that ToM is not recruited during scalar implicature derivation in children might indicate that children rely on different cognitive skills while developing the ability to derive scalar implicature.

There is evidence that adults adapt to a speaker's epistemic state in studies measuring online comprehension (Breheny et al., 2013) and reading times of scalar interpretation (Bergen & Grodner, 2012). Additionally, in a recent study by Fairchild and Papafragou (2021), ToM was found to predict scalar implicature ability in adults; in fact, EF, when controlled for ToM, did not. Moreover, there is evidence that even preschool-aged children make sensitive adaptations to a speaker's epistemic state during scalar inference (Kampa & Papafragou, 2020, 2023; Papafragou et al., 2018). However, in other studies, researchers have often failed to find a correlation between ToM tasks and pragmatic tasks (e.g. in adults, Antoniou et al., 2013; see Matthews et al., 2018 for a review) or have pointed to discrepancies between ToM and scalar implicature abilities, leading some to claim that young children compute scalar implicatures in the absence of epistemic reasoning (Barner et al., 2018, 2016; Hochstein et al., 2018). The mixed results reported in the literature may in part be attributed to differences in study design; ToM and scalar implicature tasks vary, and many studies have focused on the relation between ToM and scalar implicature reasoning in adults, making it hard to draw clear conclusions about the development of these skills.

Additionally, individuals with Autism Spectrum Disorder, a neurodevelopmental disorder whose symptoms include difficulties in social interaction that have been linked to ToM deficits (although see Gernsbacher and Yergeau [2019] for a review of empirical evidence that does not support the claim that autism is associated with an impaired ToM), are still able to compute scalar implicatures, although, in many of the studies showing success, ToM ability was either not measured or reported to be relatively high (Chevallier et al., 2010; Hochstein et al., 2018; Pijnacker et al., 2009; Van Tiel & Kissine, 2018; although see also Noveck et al., 2007). Few studies have tested scalar implicatures, ToM, language, and EF with the same participants to compare these variables directly.

Resolving the role of various supporting mechanisms in children's development of scalar implicatures is critical to the pragmatics literature; children provide a unique opportunity to gain insight into the mechanisms underlying scalar implicature derivation that adults cannot provide. Unlike adults, who have fully-developed EF, ToM skills, and language abilities, children in their preschool years are undergoing development in all three of these areas, as well as in the development of scalar implicatures. Studying children in this age group can provide insight into the extent to which these abilities develop in tandem but independently or, alternatively, scaffold each other (Saxe et al., 2006).

We propose a novel approach to investigating the mechanisms employed during preschool children's derivation of scalar implicatures: spatial neuroimaging. Behavioural research on scalar implicature, language, EF, and ToM development can only go so far in evaluating the contributions of these skills. Behavioural research can identify a correlation between two abilities (e.g. scalar implicature derivation and ToM), but cannot assess whether the same cognitive processes underpin both abilities. Similarly, behavioural research can reveal similar results between groups (e.g. children and adults), but cannot assess whether both groups arrive at the same behavioural performance using the same cognitive processes. On the other hand, by comparing spatial patterns of neural activation during scalar inferencing, neuroimaging methods provide useful insights into the cognitive processes (e.g. ToM) that are engaged during a task (e.g. scalar inference), and can reveal differences in children's and adults' cognitive processing that cannot be observed with behavioural measures alone. Behavioural studies are also limited by the types of responses that young children can give and understand, with criticism if they are too metalinguistic (Katsos & Bishop, 2011) or too short (Apperly, 2012; Matthews et al., 2018).

Neural systems and scalar implicature

Neuroimaging studies of scalar implicature and other types of pragmatic reasoning in adults have indicated the recruitment of frontal areas such as the left inferior frontal gyrus (LIFG), a region associated with higher-level linguistic processing, as well as the left anterior Middle Frontal Gyrus (MFG) and the Medial Frontal Gyrus (mFG), two areas associated with executive functions (Politzer-Ahles & Gwilliams, 2015; Shetreet et al., 2014a). To our knowledge, one study has studied the neural activation of scalar implicature derivation in children using spatial neuroimaging methods (Shetreet et al., 2014b). In this study, 6-year-olds viewed a scene and heard a sentence describing it; the sentence contained a quantifier (*some*, *none*, *all*) and either matched the picture or did not (e.g. "Some elephants are dancing" when the picture either depicted SOME or ALL of the animals dancing). Like adults, 6-year-olds displayed activation in the LIFG in contexts that had the quantifier *some*. Nonetheless, it is valuable to study children in the critical stages of development for scalar implicatures (ages 3 to 5) and the abilities that support this development. There have been additional studies on online scalar implicature derivation using EEG/ERP (electroencephalography/event-related potential) recording in both adults and children (Hartshorne et al., 2015; Noveck & Posada, 2003; Politzer-Ahles et al., 2013; Spychalska et al., 2016, 2021); while useful, these studies cannot provide the spatial information required to identify activation of neural networks associated with cognitive processes (ToM, EF, etc.), which may be active during scalar implicature.

Concerning the relationship between scalar inference and ToM, recruitment of the dorsomedial prefrontal cortex (dmPFC) and the right temporo-parietal junction (TPJ), which are involved in perspective taking and ToM, has been indicated during communicative processes (Sassa et al., 2007; Willems & Varley, 2010), although not specifically with scalar implicature derivation. For ToM more broadly, spatial neuroimaging studies investigating children's competency are very limited in number. An fMRI study by Gweon et al. (2012) investigated the activation in neural regions associated with ToM in children between the ages of 5 and 11 while children listened to stories that had "mental" descriptions (e.g. "Mary thought that Sarah … "), "physical" descriptions ("Mary walked to the store … "), or "social" descriptions (e.g. "Mary played the flute, and everybody at the party danced … "). Children displayed comparable activation to adults in bilateral TPJ, prefrontal cortex, and the medial prefrontal cortex (mPFC) for the mental stories compared to the physical stories. Additionally, participants displayed increasing selectivity in bilateral TPJ to mental state information with age; children with more advanced development of ToM displayed more specialised recruitment of neural regions associated with ToM.

There are considerably more neuroimaging studies investigating EF in children; the literature becomes extensive beyond the age of 7 (Cantlon et al., 2006; Ciesielski et al., 2006; Dibbets et al., 2006; O'Hare et al., 2008; see Houdé et al., 2010 for a meta-analysis). Studies using adolescent participants have found activation in the prefrontal cortex, consistent with the neuroimaging literature on EF in adults (Carrion et al., 2008; Crone et al., 2008; Olesen et al., 2006; O'Hare et al., 2008; Schulz et al., 2004). Children also display recruitment of frontal networks, but with less consistency and in more limited areas (Bunge et al., 2002; Crone et al., 2006a, 2006b; Crone et al., 2008; Dibbets et al., 2006; O'Hare et al., 2008), for example, activation for working memory in 7- to 9-year-olds specifically in the left ventral prefrontal cortex rather than other frontal and parietal regions (O'Hare et al., 2008).

Lastly, neuroimaging studies of language processing in young children find activation in a largely-left lateralised network involving inferior frontal, temporal, and parietal regions (e.g. in 4-year-old children, Jasińska et al., 2021). A meta-analysis of developmental neuroimaging research of language networks reported an age-related increase in neural activation in these areas (Weiss-Croft & Baldeweg, 2015). In particular, neural activation in the left inferior frontal gyrus (LIFG), which is associated with lexical access and syntatic processing (e.g. in school-aged children, Jasińska & Petitto, 2013; in adults, Price, 2012), as well as the middle temporal gyrus (MTG), which is associated with semantic processing (e.g. in adults, Hickok & Poeppel, 2004), is likely relevant to the semantic and syntatic demands associated with scalar implicature reasoning. There is evidence that connections between these areas increase with age (Bitan et al., 2006), and that the language network is largely developed as early as age 6 years (Croft et al., 2014; Wilke et al., 2009), albeit the literature on children between the ages of 3 and 5 remains sparse.

In sum, existing neuroscience literature on the development of scalar implicature, ToM, EF and language systems studies provides insight into the abilities of older children. However, research is needed on children who are still in the primary developmental stages of these cognitive areas (ages 3 to 5). Work with this target age range, however, has been limited by the challenges of using fMRI methods, which can be intimidating and too restrictive of movement for young children. This problem can be resolved by using functional Near Infrared Spectroscopy (fNIRS) neuroimaging (Fishburn et al., 2014; Gallagher et al., 2012; Jasińska et al., 2021; Nishiyori, 2016; Quaresima et al., 2012; Walsh et al., 2017). fNIRS is a neuroimaging method that utilises light absorption and reflection to measure hemodynamic activity. This method is child-friendly because it can tolerate movement better relative to MRI, and it is less intimidating for young children to wear a cap on their head than to be placed in an MRI chamber (Gallagher et al., 2012; Nishiyori, 2016; Quaresima et al., 2012).

Current study

The primary focus of the present study was to explore the neural networks active during scalar implicature derivation for both 3.5- to 5-year-old children and adults. We used behavioural and neuroimaging (fNIRS) methods to investigate the contributions of ToM, EF, and language systems to scalar implicature derivation during development and adulthood. Child participants completed a behavioural individual differences task battery (with ToM, EF, and language tasks) as well as two neuroimaging tasks using fNIRS; adult participants completed only the neuroimaging tasks. The neuroimaging tasks consisted of an ad-hoc scalar implicature task and a ToM task to directly compare neural activation for mentalizing to activation in the scalar implicature task.

Based on the results of past findings (Kampa & Papafragou, 2020), we predicted that pragmatically-sophisticated children would differentiate between more and less informative statements in the scalar implicature task in accordance with the expectation that a speaker should be as informative as their knowledge allows them to be. Of interest was whether children's performance on the language, ToM, and EF components of the individual differences behavioural battery would predict their behavioural performance on the scalar implicature task.

We compared neural activation in children, who were still developing pragmatic abilities, to adults when completing the same scalar implicature and ToM tasks. Of particular theoretical interest was the contribution of brain networks associated with ToM to scalar implicature derivation, given the diverging predictions of prominent pragmatic theories. With respect to ToM, given previous results (Gweon et al., 2012), we expected that adult participants would display activation in neural regions associated with ToM (e.g. dmPFC).[1] Given that 3.5- to 5-year-old children are still developing ToM, (Gweon et al., 2012; Saxe et al., 2009), we were interested to see whether children showed selective activation of the regions previously associated with ToM and active in adults during the ToM task (i.e. dmPFC, TPJ) during mentalizing, and whether children's neural activation would correspond with their scalar inference development. If ToM is involved in the scalar inference derivation process, we would expect to find activation of neural networks associated with ToM, including activation in the dmPFC. If there is a relationship between mentalizing and scalar inference, we expected to find at least some overlap in the neural activity for these two processes.

To our knowledge, this is the first time pragmatic inference in children between the ages of 3.5 and 5 years has been studied using spatial neuroimaging techniques. The novel approach of using neuroimaging methods to address questions raised in previous behavioural research provides a unique perspective into existing debates on the nature of pragmatic development.

Methods

Participants

Thirty-four preschool children ($M_{age} = 4;10$, *range:* 3;9 to 6;3, 19 female) and 26 adults (undergraduate students; 20 female) participated. An additional three children who participated dropped out of the study before completing the scalar implicature task and were excluded from analysis. Children were recruited from the Newark (DE) area through local postings and targeted ads on Facebook. They received a small gift for participating, and their parents received monetary compensation for their travel. Adult participants were undergraduate students at the University of Delaware and received course credit for their participation in the study. All participants were right-handed, monolingual English speakers.

G*Power software, version 3.1.9.6, was used to compute power (Faul et al., 2009). Power calculations assuming $\alpha = .05$, power $(1-\beta) = .8$, one predictor (condition), indicate a sample of 34 is sufficient to detect an effect size of 0.25, which is considered small (Cohen, 1992). A recent review of fNIRS studies found that the median sample size of published studies (included in the review) was 20 (Kohl et al., 2020).

We collected socioeconomic status, demographic, and language background information for our child participants. The majority of children were white and had college-educated mothers. Parents completed an SES ladder survey (MacArthur Scale of Subjective Social Status; Adler et al., 2000) in which they reported how they viewed their standing (income, education, respect) compared to 1) other members of their community, and 2) other people in the United States of America. Our participants' parents reported a slightly higher than average SES ladder score (*Range:* 0–10; $M_{sample} = 6.5$, $SD_{sample} = 1.77$; $M_{USA} = 6.13$, $SD_{USA} = 1.31$).

Materials

Behavioural individual differences battery (children)

For child participants, the study began with a short play period, intended to help children feel more comfortable in the lab space, followed by a behavioural test battery, which consisted of a ToM task, a vocabulary task, and an EF task in that order (the last two tasks were completed from the NIH Toolbox Cognition Battery using the NIH Toolbox iPad app; Weintraub et al., 2013). The ToM task was the classic Sally-Anne task (Baron-Cohen et al., 1985), in which children were told a story where a character has a false belief about where an item was located, and the child was asked to indicate where the character would look for that item. Participants next completed the Picture Vocabulary Test (Gershon et al., 2014). In this 5-minute task, children were presented with four images on a screen, heard an audio-recording of a word, and had to select which image the word was describing; one child did not finish this task. The EF task was the Dimensional Change Card Sorting task (DCCS; Zelazo et al., 1996). In this 10 minute task, children had to learn to match pictures based on colour, then switch to matching them based on shape, then match them either based on colour or shape after hearing an audio-recording of either the word "colour" or "shape" before each trial. The DCCS task has been shown to have high convergent validity with other EF tasks (Zelazo et al., 2014), such as the Flanker task (Eriksen & Eriksen, 1974), as well as more comprehensive measures of EF, such as the Delis-Kaplan Executive Function System (D-KEFS; Delis et al., 2001). Four children lost patience with the length of the behavioural battery and did not complete the final DCCS task but were still included in analyses for the tasks they did complete.

Neuroimaging tasks (adults and children)

Participants completed two neuroimaging tasks: a scalar implicature task, adapted for neuroimaging (Kampa & Papafragou, 2020), and a ToM task (Gweon et al., 2012). Children always performed the tasks in this order, in the interest of prioritising the scalar implicature task if children did not complete both measures ($N = 4$). The order of tasks for adults was counterbalanced. Participant responses in the neuroimaging tasks were collected using a button-box (Cedrus Inc). Prior to beginning the neuroimaging tasks, child participants completed an introductory phase in order to become familiar with the types of displays used in the test trials for the scalar implicature task. Children were also trained on how to use a button-box to indicate their answers (adult participants simply received written instructions).

Scalar implicature task. The scalar implicature task was adapted from a paradigm used previously to test 4- and 5-year-olds' scalar inference derivation (Kampa & Papafragou, 2020). For this task, participants viewed two photos placed side by side on a computer screen (see Figure 1 for an example). Each photo depicted a person ("the experimenter's friend") sitting across from and facing out towards the participant (for example, the girl). In front of her was a cardboard box with two vertical compartments. In the "full-access" box photo, both vertical compartments of the box were open and see-through, thus the participant and the girl both had full visual access to the box (photo on the left within each image in Figure 1). In the "limited-access" box photo, the compartment to the left (from the participant's viewpoint) was see-through, like the compartments in the full-access box (photo on the right within each image in Figure 1). However, the compartment to the right was blocked such that only the participant but not the girl could see its contents. The contents of the two boxes were identical (e.g. a penguin and a pumpkin), although one of the objects in the limited-access box was not visible to the experimenter's friend (here, the pumpkin; Figure 1). After both photos were presented, children heard a recorded sentence from a female speaker ("I see … ") and were asked, "Which box is she talking about?" and had to select one of the photos.

There were two target within-subjects conditions, Perspective-Taking and Scalar-Inference, and a Control condition (the first two conditions correspond to the conditions named More-Informative and Less-Informative, respectively, in Kampa & Papafragou, 2020). The Perspective-Taking and Scalar-Inference conditions were designed to be minimally different from each other. Both conditions used the same photos (counterbalanced between participants) and required the participant to evaluate the visual perspective of the speaker. The target Scalar-Inference condition differed from the Perspective-Taking condition only in that the speaker's statement provided less information ("I see a penguin" vs. "I see a penguin and a pumpkin" respectively). A successful responder in the Perspective-Taking condition should recognise that only the speaker with full visual access could produce the more informative statement ("I see a penguin and a pumpkin"), since the speaker with limited knowledge could not see one of the objects; the responder should thus take the full-access box to be what the speaker is talking about.

A successful responder in the Scalar-Inference condition should recognise that a communicator with full knowledge should not fail to present complete and relevant information (i.e. the full contents of the box), and thus match the less informative statement ("I see a penguin") to the speaker with limited knowledge, hence picking the limited-access box as the one being talked about (see Papafragou et al., 2018 or Kampa & Papafragou, 2020 for a more detailed explanation of this logic). The control condition was designed so participants could make the correct selection without needing to rely on perspective taking or scalar reasoning abilities; the child only needed to identify the picture with the object (e.g. penguin vs. banana) matching the heard utterance ("I see a penguin"). Control blocks also ensured that children could select an image that matched the utterance they heard; low performance on scalar implicature reasoning and perspective conditions would therefore not be attributed to a child's inability to complete the simpler control task.

The task used a block-design.[2] Participants completed 2 runs lasting approximately 5 minutes each. For children, each run consisted of 6 blocks of test trials (2 Perspective-Taking, 2 Scalar-Inference, 2 Control blocks) in a counterbalanced order. Adults completed runs with 2 perspective-taking and 2 scalar inference blocks, but did not complete a control block. Each block contained 4 trials (12 seconds each) of the same type. For each trial, the photo on the left appeared on screen first, followed by the photo on the right after 2 seconds (presentation order for the limited access box photo was counterbalanced across participants). After both photos were presented, children heard a recorded sentence from a speaker matched to the person in the photo ("I see … "), then a recording from a distinct narrator voice

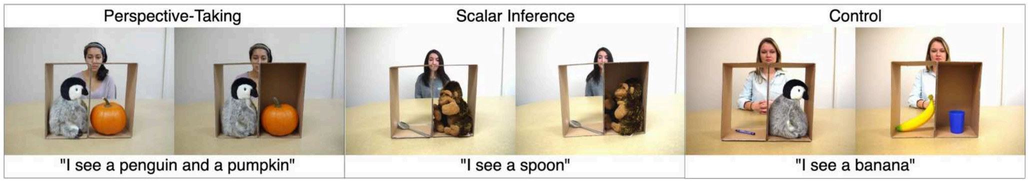

Figure 1. Example Trial from Each Condition.
Note: The expected box selection for each condition is highlighted in the corresponding color.

asking, "Which box is she talking about?" Participants were given 6 seconds after the audio ended to indicate their response by pressing the appropriate button; both photos remained on screen for those 6 seconds. For variety, the people in the pictures (and their corresponding recorded voices) differed across blocks; each "friend" was depicted in two blocks total (always different types) across the experiment. The interstimulus interval (ISI) between each block was 15 seconds long; adults saw a black screen with a fixation cross, children saw a video of jellyfish moving (Run 1) or small, colourful dots moving (Run 2).

ToM task. Participants completed a neuroimaging ToM task adapted from a study by Gweon et al. (2012). There were two within-subjects conditions: Mental and Physical. In the Mental condition, participants listened to stories that described a protagonist's mental state using mental state verbs; in the Physical condition, participants listened to stories that described physical events, objects, and states (see Table 1). The stories were 20 seconds in length and matched for the number of words ($M = 51.6$ words), number of sentences (4.7), and Flesch Reading Ease Level ($M = 90.4$). After each story, participants were asked, "Does this come next?" followed by a probe sentence that either matched the story or came from an unrelated story (4 seconds). Participants indicated their judgment by pressing either a green button for "yes" or a red button for "no" (6 seconds). The question was designed to verify attention; of interest was the neural activation while listening to the two types of stories, not the participants' behavioural responses.

Table 1. Example stimuli for fNIRS ToM task.

Condition	Example story	Example probe
Mental	"One day a pirate told Jimmy about a hidden treasure. The pirate **thought** that the treasure was buried behind Jimmy's house. Jimmy **believed** him. So Jimmy dug a big hole behind his house, but he didn't find a treasure. Jimmy soon **realized** the pirate didn't **know** where the treasure was."	"Jimmy was mad. He had done all that work for nothing."
Physical	"One day, a little robin landed on a huge, strange-looking tree, and ate one of its berries. This was a magical tree that had special powers. In the spring, the robin laid three eggs. Soon, two of the eggs cracked and little robins came out. But the last egg did not crack for a long time."	"When the last egg cracked a beautiful firebird came out."

Note: For a full list of stimuli, see the supplementary materials for Gweon et al. (2012).

Each run consisted of 4 trials in a palindromic, counterbalanced order (Mental-Physical-Physical-Mental and vice versa) with a 12-second rest before each trial and lasted approximately 2 minutes and 48 seconds. Adults completed 4 runs; children were only able to consistently complete 2 runs.

Procedure

During the fNIRS testing, the hemodynamic response was measured with a Shimadzu LightNIRS Near Infrared Spectroscopy system acquiring data at 7.4 Hz. The lasers and detectors were arranged in a 3×10 array resulting in 47 channels (see Figure 2). This array was chosen to allow for coverage of frontal regions previously indicated to be active during pragmatic inference (e.g. LIFG) as well as to allow for the most general coverage of relevant EF, ToM (e.g. dmPFC[3]), and language areas possible given that fNIRS has restrictions on the amount of total surface area that can be covered by the array. The lasers were factory set to 780, 805, and 830 nm. The 15 lasers and 15 detectors were segregated into alternating grid placement. Positioning of the array was accomplished using the 10–20 system (Jasper, 1961) to maximally overlay the key regions of interest (for additional details, see Jasińska et al., 2017; Jasińska & Petitto, 2013, 2014; Shalinsky et al., 2009). The depth of recording in the cortex is approximately 3 cm. Prior to recording, every channel was tested for optimal signal to noise ratio using LightNIRS inbuilt software.

Like fMRI, fNIRS measures the brain's hemodynamic response. fNIRS has several advantages compared to fMRI; chiefly fNIRS has greater ease of use, particularly with younger children: children are

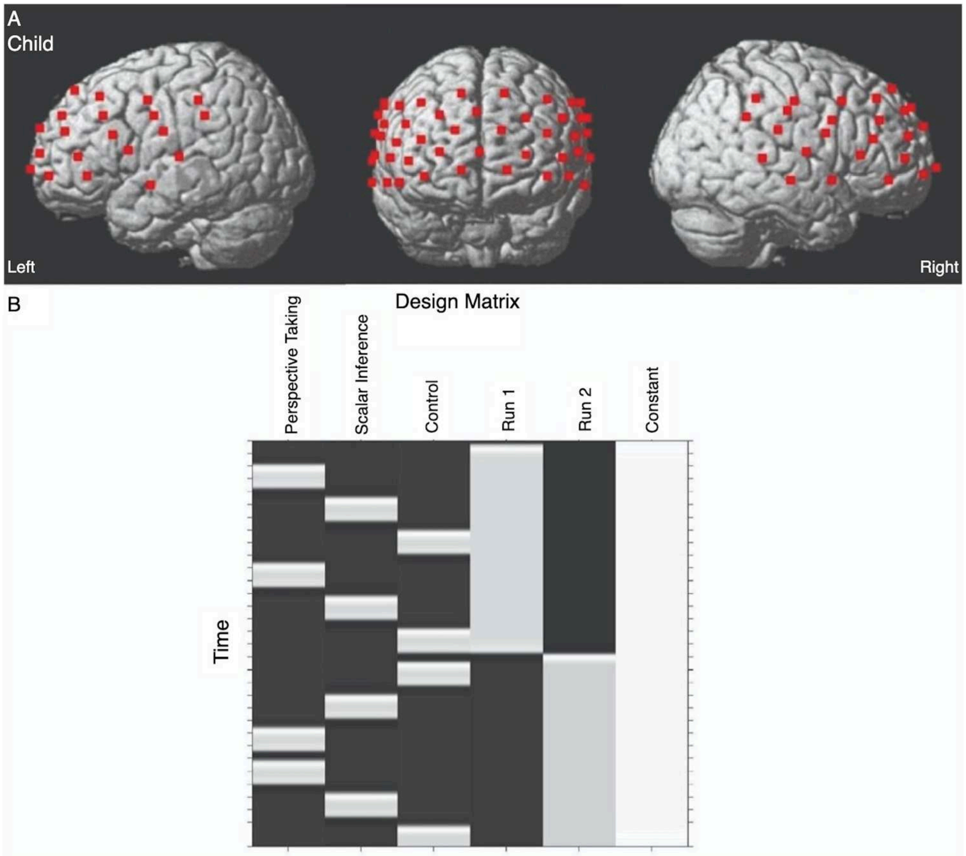

Figure 2. Arrangement of the 3 × 10 array on child participants and NIRS design matrix.

Note: (A) The probe array visualized is an average of all child participants. The probe array extended more posteriorly on the head of a child compared to the adult given the head size differences between 3.5- and 5-year old children and adults. (B) Design matrix.

seated during the session and fNIRS tolerates movement better than fMRI. However, fNIRS has poorer signal-to-noise ratio and coarser spatial resolution compared to fMRI. Importantly, fMRI BOLD and fNIRS HbO and HbR signals typically show high correspondence. Studies that examine neural responses during concurrent fNIRS-fMRI imaging sessions show moderate spatial correlations (e.g. $R = 0.57$; Huppert et al., 2017; $R = 0.26$ Cui et al., 2011; $R = 0.576–0.801$; Pereira et al., 2023). Moreover, fNIRS-fMRI correlations are not affected by task type; fNIRS-fMRI correlations are similarly high across cognitive (i.e. Go-no-go inhibitory control task, N-back working memory task), motor (i.e. finger tapping), and visual (i.e. judging line orientation) tasks (Cui et al., 2011). The majority of spatial differences between fNIRS and fMRI are driven by the lower sensitivity for deeper brain structures of fNIRS compared to fMRI.

Data analysis

Behavioural analysis. Behavioural data were analysed using R (Version 4.0.0) using the package lme4 (Version 1.10; Bates, Maechler, Bolker, and Walker, 2015). Given that the behavioural study from which this paradigm was adapted found a bimodal distribution of responses, with children either succeeding or failing across the board (Kampa & Papafragou, 2020), we used a data-driven approach (k-means cluster; Wagstaff et al., 2001) to analyse behavioural scalar implicature results and determine if there were distinct patterns of responding (e.g. pass vs. fail). To analyse the relationship between the behavioural tasks and performance on the scalar implicature task, we first conducted a set of basic analyses (two-sample *t*-tests for scale data; Fisher's exact test for categorical data) to compare the passers and non-passers. To assess whether performance in any of the behavioural tasks predicted behavioural scalar implicature task performance, the data were analysed using multi-level logistic mixed-effects modelling. Accuracy in the target condition (Scalar-Inference) of the scalar implicature task was analysed with a model that included Age (continuous), Language (Vocabulary), EF (DCCS), and ToM (Sally-Anne) scores as fixed predictors and random intercepts for subject and item. Interactions that did not significantly improve model fit were removed from the final model.

fNIRS preprocessing. Neuroimaging data were preprocessed and analysed using a Matlab-based NIRS-SPM Version 4 (Jang et al., 2009; Ye et al., 2009), which uses the neuroimaging suite SPM12 (Flandin & Novak, 2020; Nichols, 2013). Using the modified Beer–Lambert equation, NIRS-SPM converts optical density values into concentration changes in oxygenated and deoxygenated hemoglobin response (HbO and HbR, respectively). Changes in HbO and HbR concentrations were filtered with a HRF filter and decomposed using a Wavelet-Minimum Description Length (MDL) detrending algorithm in order to remove global trends resulting from breathing, blood pressure variation, vasomotion, or participant movement artifacts and improve the signal-to-noise ratio (Jang et al., 2009). We filtered out step functions that were identified in each timeseries. This allowed us to correct for motion artifacts as well as drift in signal related to respiration.

Spatial registration. NIRS channels were registered to MNI space with the Haskins Pediatric Brain Atlas (Molfese et al., 2021) in NIRS-SPM's stand-alone registration function (Singh et al., 2005) by using a three-dimensional digitizer (Polhemus Corp.). Registration was done individually for each child. The spatial registration function yielded MNI coordinates represented by each channel with corresponding labels for anatomical regions, including Brodmann labels, maximally located at each channel position. Specifically, the function provides a coverage percentage for a given anatomical region at each channel.

fNIRS analysis. We used a general linear model analysis approach that allows for the creation of activation maps with super-resolution localisation and applied Sun's tube correction for multiple comparisons. Hemodynamic response function smoothing was used for low-pass filtering. The temporal correlations were estimated using precoloring (Worsley & Friston, 1995). Models for HbO and HbR contain experimental regressors convolved with the corresponding hemodynamic response function with time derivatives. Figure 2(B) shows the design matrix. NIRS-SPM (Statistical Parametric Mapping) creates the models for HbO and HbR with opposing polarity so that a significant model fit for HbO indicates increased concentration and for HbR decreased concentration.

Scalar implicature. Group activation maps were generated comparing the target Scalar-Inference condition to the Perspective-Taking condition. Note that these conditions were carefully matched in terms of perspective taking and visual detail such that removing activation for the Perspective-Taking condition from the Scalar-Inference condition (i.e. through standard neuroimaging subtraction analysis) yielded activation specific only to the process of scalar inference. This task was specifically chosen so that any activation present after the subtraction could not be attributed to perspective taking or other mentalizing processes independent of the scalar inference. The control condition did not produce any results of note and will not be discussed further in this manuscript.

ToM. Group activation maps were generated for "mentalizing" (equivalent to the subtraction of the Physical condition from the Mental condition in this task).

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow JARS (Kazak, 2018). All data, analysis code, and research materials are available at https://osf.io/w3q8r/. This study's design and its analyses were not pre-registered. This study received approval from the the University of Delaware.

Results

Behavioural results (children): individual differences and scalar implicature performance

We identified two distinct groups of children: 20 scalar implicature passers (who achieved at least 75% accuracy in the Scalar Inference Condition; $M_{passers} = 98\%$, *range:* 87% to 100%; $M_{age} = 4.85$) and 15 scalar implicature non-passers ($M_{non\text{-}passers} = 39.5\%$, *range:* 0% to 73.8%; $M_{age} = 4.93$).[4] Children who were designated Passers in the Scalar-Inference condition performed significantly better ($t\ (33) = 6$, $p < .001$) on the Perspective-Taking trials ($M = 95.9\%$, *range:* 71.4% to 100%;), while children designated Non-Passers had more difficulty ($M = 66.8\%$, *range:* 18.8% to 100%). Results on the control conditions of the scalar implicature task are reported in Supplementary Materials. We found no significant differences between passers and non-passers in any of the behavioural battery tasks (*Language* [*Vocabulary*)]: $M_{Passers} = 63.42$, $M_{Non\text{-}passers} = 60.87$; $t(32) = 1.43$, $p = 0.16$; *EF* [*DCCS*]: $M_{Passers} = 56.71$, $M_{Non\text{-}passers} = 53.07$; $t(29) = 0.56$, $p = 0.58$; *ToM* [*Sally-Anne*]: Fisher's exact test, $p = 0.56$; see Figure 3).

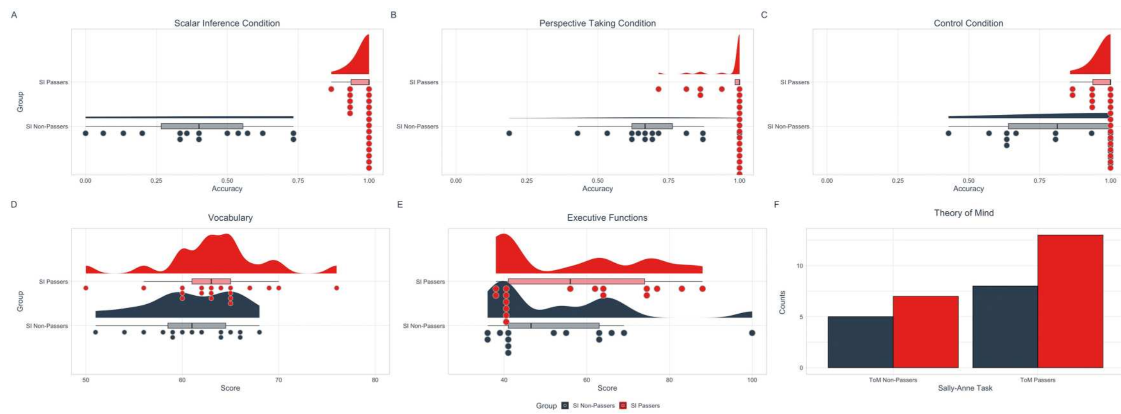

Figure 3. Performance in the behavioural individual differences battery tasks for passer and non-passer scalar implicature groups.
Note: SI Passers and SI Non-Passers indicates children who passed or did not pass the Scalar Implicature task, respectively. ToM Passers and ToM Non-Passers indicates children who passed or did not pass the Sally-Ann Theory of Mind task, respectively.

Our statistical model revealed no significant effects of age, language, executive function and ToM or interactions (see Table 2). The finding that behavioural individual differences measures failed to predict children's scalar implicature performance is reminiscent of previous studies (see Matthews et al., 2018), and motivates the need to look at the neural signatures of scalar inference derivation to gain additional evidence for the mechanisms underlying this pragmatic phenomenon.

Neuroimaging results (adults and children)

Scalar implicature

Beginning with the adult data, regions that were found to have significant HbO activation during scalar inference for adults are shown in Table 3 and Figure 4. Regions of note include the LIFG (specifically, Broca's area), dlPFC, and dmPFC. Other regions of activation include Wernicke's area, the primary somatosensory cortex, and the pre-motor and supplementary motor cortex.

While adults showed canonical patterns of HbO and HbR concentrations (increased HbO concentrations relative to HbR), children showed the inverse patterns (increased HbR relative to HbO). This phenomenon has been observed frequently with infants and very young children; evidence suggests that more complex stimuli and cognitive processes trigger this inverse shape of HbO and HbR responses (Grossmann, Oberecker, et al., 2010; Grossmann, Parise, et al., 2010; Issard & Gervain, 2017; Taga, Asakawa, Hirasawa, et al., 2003; Taga, Asakawa, Maki, et al., 2003; Watanabe et al., 2012; see Issard & Gervain, 2018 for a review). It has been proposed that children display inverse HbO and HbR patterns when stimuli are particularly difficult for the tested developmental stage (Issard & Gervain, 2018). Typically, inverse responses give way to canonical responses over the course of development (Emberson et al., 2017) as children develop more familiarity with and facility for a given ability.

For children, nine passers and nine non-passers had usable fNIRS data for the scalar implicature task. Overall, the child passer group showed only modest activation in the dlPFC during scalar implicature derivation. Next, we conducted an exploratory cluster analysis (e.g., Bonomini et al., 2015) to examine children who succeed behaviourally on the scalar implicature task. Although behaviourally identical,[5] within the passer group, there were two distinct subgroups: a subset of children who demonstrated similar patterns of neural activation to adults during scalar implicature derivation (dmPFC and dlPFC) and a subset of children who displayed no cohesive significant neural patterns during scalar implicature.

Child non-passers displayed the highest level of activation in the right temporoparietal region. Similar patterns of neural activation between adults, child passers, and non-passers included the dlPFC, Wernicke's area, and the pre-motor and supplementary motor cortex. However, child non-passers, in contrast to the child passer subgroup and adults, did not display significant activation in the LIFG or the dmPFC.

Table 2. Parameter estimates for behavioural accuracy in the scalar implicature task.

Effects	Estimate	SE	*z*
Intercept	6.010	5.235	1.148
Age (continuous)	−.824	1.064	−.775
Language	.207	.128	−1.614
Executive functions	.028	.032	.884
Theory of Mind	.508	1.136	.448
Age (continuous) × language × Theory of Mind	.036	.050	.722
Language × executive functions × Theory of Mind	−.014	.015	−.923

Note: Significance levels: *$p < .05$, **$p < .01$, ***$p < .001$.

Table 3. Regions with significant activation for scalar inference (scalar inference > perspective taking) and theory of mind (mental > physical).

Region	Brodmann area	MNI coordinates X	Y	Z	β
Scalar implicature (scalar inference > perspective taking)					
Adults (HbO)					
Broca's area	45	−16.3	52.3	44.3	.00259
R. primary somatosensory cortex	1/2/3	64.7	−19	46	.0024
dlPFC	9	−47.3	4.3	55.7	.0023
dmPFC	9	−10.3	40.3	56.7	.0022
R. pre-motor and supplementary motor cortex	6	64.6	−19	46	.0019
L. pre-motor and supplementary motor cortex	6	−53	9.7	44.7	.0017
dmPFC	9	−16.3	52.3	44.3	.0016
dlPFC	46	49	53.3	−5.7	.0016
Wernicke's area	40	−47.3	−37.3	63.3	.0015
L. frontopolar area	10	−13.7	64.7	29.3	.0015
Children Passers Subgroup (HbR)					
Broca's area	45	50	47.7	17.7	.0034
dlPFC	9	41	36	44	.0021
dmPFC	45	−2.3	58.7	38.7	.0017
L. primary somatosensory cortex	11	−69	−30.3	23.7	.0016
Children Non-Passers (HbR)					
R. superior temporal gyrus	22	71	−34.3	23.7	.0034
R. middle temporal gyrus	21	73	−26.3	−7.7	.0029
dlPFC	46	46	50.7	22.7	.0025
R. supramarginal gyrus	40	68	−42.3	32.7	.0025
L. pre-motor and supplementary motor cortex	6	−64	4.3	25.3	.0022
L. frontopolar area	10	−11.3	74	5.3	.0020
Theory of mind (mental > physical)					
Adults (HbO)					
LIFG	44/45	−54	24.6	29.6	.0016
R. premotor & supplementary motor cortex, R. DLPFC	6/9	47.7	15.7	53.7	.0011
R. dlPFC, RIFG	9/45/46	42.3	39.3	40.3	.0010
Children Passers (HbR)					
dmPFC	9	25.6	43.6	48.6	.0006
R. supramarginal gyrus	40	70	−38.6	24.3	.0005

Note: For adults, this table displays the ten channels with the highest level of significance; the rest can be found in supplementary materials. *p*-value is less than .05 for all contrasts. HbR (adults) and HbO (children) are included in Supplementary Materials.

We compared the passer and non-passer groups, as well as the passer subgroup to the non-passer group on the perspective-taking condition and did not find any significant differences at the neural level. Although, we are cautious about interpreting null effects, particularly with our small sample.

ToM

For adults, greater HbO activation was observed in the dlPFC, the supramarginal gyrus, and the primary somatosensory cortex (Figure 4, Table 3). However, there was no significant HbO activation in the dmPFC during mentalizing, even though the stimuli were identical to those used in a previous fMRI study that found activation in the dmPFC (Gweon et al., 2012).[6] To assess the relationship between scalar inference derivation and ToM, we compared patterns of neural activation during the ToM task for children in the passers and non-passers. The child passer group displayed recruitment of the dmPFC and right supramarginal gyrus (part of the right TPJ) during mentalizing in this task (Table 3). Child non-passers in the scalar implicature task did not display any significant group patterns of activation. Children who did well behaviourally on the scalar implicatures task (passers) displayed activation in the dmPFC and right TPJ, while child non-passers did not. This could suggest a connection between more developed ToM abilities (indicated by selective recruitment of regions associated with ToM; see Gweon et al., 2012; Saxe et al., 2009) with success on the scalar implicature task.

Discussion

The cognitive signature of scalar inference

In conversation, when speakers provide statements that give limited information ("Some toys on the table are free"), listeners typically infer that the speaker is not in a position to offer a more-informative statement ("All toys on the table are free"). This type of inference, a *scalar implicature*, has been extensively studied, yet debates continue over the contributions of various mechanisms to scalar implicature derivation and the way those mechanisms interact during development. Using a novel fNIRS neuroimaging approach combined with more familiar behavioural methods, we investigated the contributions of cognitive processes including ToM, EF, and language to the derivation of scalar implicatures during development as well as in adulthood.

Behavioural measures of ToM, EF, and language failed to predict children's behavioural scalar inference performance in our study (consistent with previous studies; see Matthews et al., 2018 for a review). In contrast, previous work by Fairchild and Papafragou (2021) found that ToM predicted scalar implicature ability; although there are key differences between the present study and Fairchild and Papafragou's work. Fairchild and Papafragou (2021) focused on a larger sample ($n = 178$) of adults who completed different ToM ("Mind in the Eyes Task"; Baron-Cohen et al., 1997, 2001; "Strange Stories Task"; Happé, 1994) and scalar implicature (involving ratings of informative and underinformative sentences) tasks (see Fairchild & Papafragou, 2021, for task details).

fNIRS technology allowed us to directly examine neural activation during scalar inference to gain insight

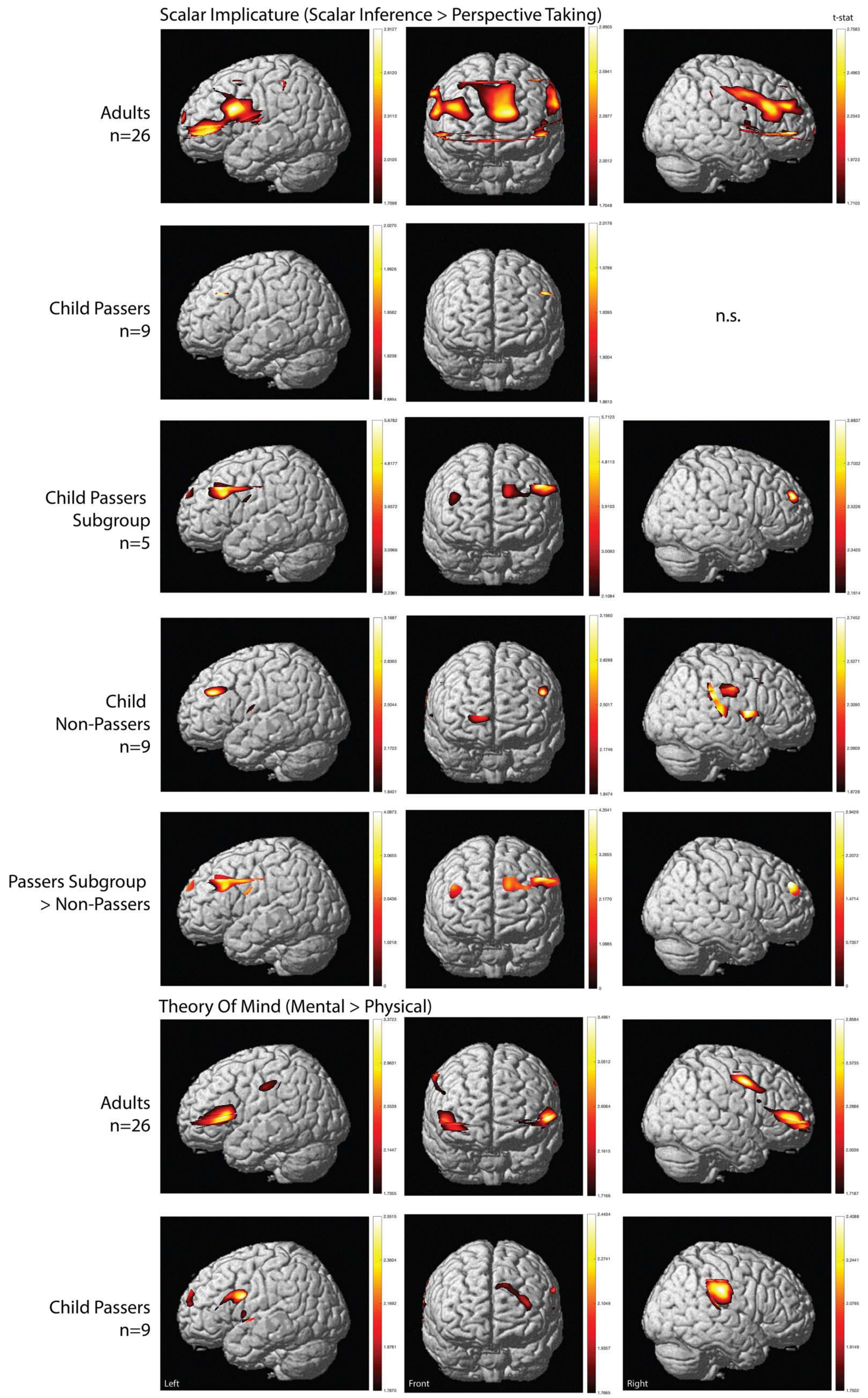

Figure 4. *t*-statistic maps showing neural activation patterns in the scalar implicature and theory of mind tasks for adults, child passers, and child non-passers.

Note: Activation is shown for HbO for adults and HbR for children (HbR for adults and HbO for children can be found in supplementary materials). *p*-value is .05 for all contrasts.

into the cognitive processes involved in the derivation process, something not possible using behavioural methods alone. Our neuroimaging results indicated that neural areas associated with ToM, EF, and language networks are active during adults' scalar inference. Specifically, we found significant activation for adults in the dmPFC, dlPFC, and LIFG during scalar inference, which have been previously implicated in ToM, EF, and high-level linguistic processing. Here, we note the overlap in the networks that are active for adults during both scalar inference derivation and mentalizing. In adults, regions active during mentalizing were also active during scalar inference derivation, specifically the LIFG and dlPFC. This could indicate a connection between the processes involved in ToM reasoning and scalar inference derivation, although there is always the possibility that both processes recruit the same regions independently. While it is difficult to make definitive claims from this exploratory evidence, these results are suggestive that there may be a relationship between ToM and scalar inference.

We provide the first evidence from spatial neuroimaging techniques that some 4- to 5-year-old children display similar patterns of neural activity to adults during scalar inference, specifically activation in the dmPFC, dlPFC, and LIFG. Interestingly, although all children identified as passers on the scalar implicature task showed similarly high accuracy in the scalar implicature task, there was no clear group pattern at the neural level. Instead, despite behavioural group similarity, we found neural group variability. A subgroup of child passers showed a cohesive neural profile similar to the patterns observed for adults, while the remaining child passers did not. Inter-individual variability in neural response has previously been observed in children, with greater variance in neural activation patterns in childhood that homogenise over development. For example, studies that have examined inter-individual variability in neural response have found that infants show greater variation in neural responses for language compared to adults (Wild et al., 2017), and children show greater variation in neural representations for faces compared to adults (Tian et al., 2021). However, other research suggests that inter-individual variability may also increase over development. For example, adults showed greater variation in neural responses during passive viewing of real-world stimuli than children (Petroni et al., 2018).

While both the child passer subgroup and child non-passers in the scalar inference task displayed activation in the dlPFC, only passers displayed activation in the dmPFC and the LIFG, regions associated with ToM networks and higher-level linguistic processing networks, respectively, akin to adult participants. This could suggest a relationship between more developed ToM, cognitive, and language networks with more adult-like patterns of neural activity during scalar inference. During this developmental period, some children may be able to demonstrate scalar implicature reasoning behaviourally but may be leveraging different neural systems to do so. In adults, neural similarities may reflect an accumulation of similar experiences with scalar implicature reasoning and similar strategies for processing these constructions. The adult-like subgroup of pragmatically sophisticated children may already be processing scalar implicatures in the same manner. This evidence supports the viewpoint that at least some children are capable of employing sophisticated, adult-like pragmatic reasoning processes from a fairly young age (Clark, 1990; O'Neill, 1996; Diesendruck & Markson, 2001; Nadig & Sedivy, 2002; Matthews, et al., 2006) and suggests that the development of ToM, EF and language could be interacting with pragmatic inference during this developmental period between the ages of 3 and 5.

Interestingly, child non-passers specifically showed increased activation in the right TPJ – a region associated with ToM – during scalar inference. This activation may suggest that more effortful ToM processing is involved in reasoning about scalar implicature derivations, even during a developmental stage where children are not yet sophisticated enough to behaviourally pass the scalar implicature task. ToM may be required to process scalar implicatures, as indicated by increased right TPJ activation in non-passers, but not sufficient without more mature language and executive function skills. Moreover, children who passed the scalar implicature task, but not non-passers, showed robust right TPJ activation during the ToM task, further suggesting that ToM processing is more developed in pragmatically sophisticated children. Child non-passers who could perform well on the ToM task would presumably have sufficiently developed ToM networks to support ToM task performance. Yet, at the group level, neural patterns consistent with ToM ability were only observed in the passer group, suggesting that ToM networks may remain relatively underdeveloped or inefficiently used by non-passers during these tasks. It's important to note that here, again, we observed differences between scalar implicature passers and non-passers in neural responses during ToM processing, with no corresponding difference between groups in behavioural ToM task performance. This observation further underscores the insights afforded by neuroimaging investigations that are not captured by behavioural tools alone.

Taken together, these results provide novel evidence bearing on theoretical debates over the nature of pragmatic inference, specifically the role that ToM plays during this process. Under Gricean and several newer inferential accounts, scalar implicature derivation is a complex, social-inferential process that requires the listener to incorporate information about the speaker's knowledge state and other characteristics (Grice, 1975, 1989; Sauerland, 2004; Sperber & Wilson, 1995/1986, 2002). However, lexical and grammatical accounts have proposed that scalar implicature derivation can often be achieved through linguistic mechanisms alone without the need for epistemic reasoning (Chierchia, 2004, 2006; Chierchia et al., 2009; Fox, 2007; Levinson, 2000) and that preschool children do not necessarily engage in epistemic reasoning during scalar inference (Barner et al., 2018; Hochstein et al., 2014, 2018). This study provides evidence that neural networks activated during mental reasoning are implicated in scalar implicature computation in adults and even some young children.[7] This evidence is critical for Gricean and other inferential theories, and suggest that grammatical or lexical theories may need to include additional assumptions in order to account for the results presented here.

Limitations and future directions

Although this study provides a vital foray into the mechanisms supporting scalar inference in adults and young children, it has some limitations. The most apparent limitation is the low number of participants in the child groups. Despite our best efforts, it was challenging for 4- and 5-year-old children to complete a multi-task neuroimaging study. The children had to wear a cap with electrodes attached while trying to sit as still as possible in a dark room during the tasks (fNIRS data collection is sensitive to light and movement), separate from their parents to avoid distractions. We also faced recruitment challenges for this type of study given the relatively small local pool and COVID-related difficulties impacting in-person lab testing.

Increasing the number of participants could allow for a more detailed exploration of the differences between child passers and non-passers (focusing on individual differences). An interesting follow-up line of inquiry would be to test a wider age range with a focus on individual differences, possibly with a longitudinal component, to gain insight into how and when this change is brought about in children. The literature on developmental changes in inter-individual variability in neural response is sparse, with mixed results (e.g. an increase [Tian et al., 2021; Wild et al., 2017] and decrease [Petroni et al., 2018] in inter-individual neural response variability over development)–this is a needed area of more research. Additional work should also explore individual differences in ToM task activation, particularly concerning whether and how any developmental shifts in ToM activation patterns relate to scalar implicature reasoning. This is important to consider, given that our scalar implicature condition was contrasted with perspective-taking–understanding developmental shifts in the processes involved in this experimental condition will further inform our understanding of both ToM and scalar implicature reasoning. Additionally, it will be important for future studies to examine whether increased neural activation during scalar implicature reasoning and ToM reflect effortful processing or efficient engagement of neural substrates required for success on each task.

Other directions could include testing children (and even adults) with autism to investigate whether they display similar patterns of neural activation to non-autistic individuals like in this study, especially given that autistic individuals have shown success with scalar implicatures (Chevallier et al., 2010; Hochstein et al., 2018; Pijnacker et al., 2009; Van Tiel & Kissine, 2018) despite some studies indicating impaired ToM reasoning. However, it is important to note that empirical data supporting the claim that autistic people lack a ToM is mixed (i.e. empirical evidence fails to demonstrate impaired ToM in autism, studies showing impaired ToM in autism failed to replicate, ToM task performance does not account for autistic traits; see Gernsbacker and Yergeau [2019] for a review).

Another limitation is that, while we found activation of brain networks associated with EF (notably the dlPFC), the relationship between EF, scalar implicature, and ToM may be more complicated. In a recent study, Fairchild and Papafragou (2021) found that EF ability was correlated with scalar implicature derivation ability, but not when ToM was controlled for, albeit importantly, Fairchild and Papafraou's work focused on adults. They proposed that the correlation between EF and scalar implicatures could be accounted for by the engagement of EF during the ToM reasoning process itself (separate from the scalar implicature). This introduces the possibility that neural networks associated with EF may be active during scalar inference in our task as a by-product of the ToM reasoning involved, rather than through a direct reliance on EF for scalar implicature derivation, and it is difficult to make definitive claims based on the presence of dlPFC activation during scalar inference derivation about how that activation should be interpreted. More comprehensive measures of ToM and EF with these neuroimaging techniques might be helpful in further teasing out the contributions of EF to the scalar implicature process separately from the contributions of ToM.

This study opens a number of possible directions for continued inquiry into the development of pragmatic inference in preschool children. First, the paradigm we used relies on an ad hoc (entirely context-dependent) scale, which might increase the salience of relevant scalar alternatives through the presence of context; it would be interesting to see whether these results generalise to quantifier scales (e.g. involving *some* and *all*) within the same age group (see Barner et al., 2016; Katsos & Bishop, 2011; Papafragou & Musolino, 2003). Second, the present method explicitly contrasted a speaker with full vs. partial knowledge in situations where participants themselves always had full knowledge. Future versions of this method could test whether ToM is involved in implicature-derivation even if the speaker's and the participant's perspectives align (e.g. if the sentence "I see a penguin" has to be matched either to a speaker who sees a penguin only, or to a speaker who sees a penguin and a pumpkin, with no hidden objects from the speaker's point of view; in that case, the first choice is the pragmatically appropriate one). Additionally, it will be important to further examine shared neural substrates underlying scalar inference and perspective taking across a broader range of neuroimaging tasks that contrasts scalar inference, perspective taking, with other cognitive processes that may be common to both.

Lastly, our findings provide evidence of a relationship between ToM and language development with scalar implicature development, but the scope of this paper precludes us from making stronger claims as to whether a causal relationship exists between these systems. Future work should continue to elucidate the nature of the connections between these mechanisms.

Conclusion

Here we presented the first spatial neuroimaging study to look at scalar inference in 3.5- to 5-year-old children and one of the very few studies to study early pragmatic development using neural measures more broadly. We chose to investigate scalar implicatures because they have been extensively studied using behavioural measures; however, the present neuroimaging approach could easily be applied to other types of conversational inferences or other pragmatic phenomena that have not been as extensively explored. The regions that we found to be active during scalar inference (dmPFC, dlPFC, LIFG) in both pragmatically-sophisticated children and adults are likely to also be recruited during other types of pragmatic inference, and the role that supporting mechanisms such as ToM, EF, and language play in this process plausibly generalise to other types of pragmatic inference, at least according to Gricean models of inferential communication (see Fairchild & Papafragou, 2021). The current study provides not only a crucial look below the surface into children's scalar implicature derivation, but also a novel contribution to the broader developmental study of pragmatics and conversational implicature, bridging the gap between behavioural patterns for conversational inference and the neural substrates that underlie young learners' pragmatic calculations.

Notes

1. Neural regions covered by the fNIRS array we chose to reflect the broad scope of the cognitive abilities employed across all of the tasks and the within approximately 3 cm depth of the cortex based on the penetration of near-infrared light in tissue.
2. Block designs are commonly used in child neuroimaging protocols as they allow for the capture of the hemodynamic response generated by the experimental stimulus. Repeat trial presentation within block designs reliably engages neural regions associated with task processing; trials have typically closely spaced, successive presentations over a short interval of time (15–50s). Block designs are efficient because they allow for better signal-to-noise ratio by collapsing across many trials (e.g. Bandettini, 1993). For young children, who tend to generate movement-related artifacts, block designs, as compared to event related designs, offer the advantage of improved signal-to-noise ratio.
3. Note that other regions associated with Theory of Mind, such as the bilateral temporal parietal junction, would not be covered by the surfacearea of the fNIRS array chosen for versatility in this study for adult participants given their larger head size.
4. The results do not change if we define passers in terms of accuracy in both the Scalar Inference and the Perspective-Taking Conditions.
5. A set of simple tests (two-sample *t*-tests for scale data; Fisher's exact test for categorical data) revealed no significant differences in performance on behavioral tasks between the two subgroups (*Age:* M_{G1}= 4.71, $Range_{G1}$ = 4;1 to 5;10 ; M_{G2} = 5.12, $Range_{G2}$ = 4;2 to 6;1; *Language [Vocabulary]*: M_{G1} = 67, M_{G2} = 63; $t(7) = -0.75$, $p = 0.48$; *DCCS*: M_{G1} = 69, M_{G2} = 50; $t(7) = 1.90$, $p = 0.11$; *Theory of Mind* [*Sally-Anne]*: Fisher's exact test, $p = .41$).
6. There was significant HbR activation in the dmPFC (see supplementary materials).
7. Note that this activation during the scalar task is above and beyond that required for visual perspective taking in the same task; activation for visual perspective taking is subtracted out from the primary analysis.

Acknowledgements

The researchers thank the children and families who participated in this study, and the research assistants at the University of Delaware who supported data collection.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research is supported by NSF grant #1632849 awarded to A. Papafragou.

References

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. *Health Psychology*, *19*(6), 586–592. https://doi.org/10.1037/0278-6133.19.6.586

Antoniou, K., Cummins, C., & Katsos, N. (2016). Why only some adults reject under-informative utterances. *Journal of Pragmatics*, *99*, 78–95. https://doi.org/10.1016/j.pragma.2016.05.001

Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (2013). *Does multilingualism confer an advantage for pragmatic abilities?* Retrieved June, 2020, from http://ktisis.cut.ac.cy/handle/10488/4102

Antoniou, K., Kambanaros, M., Grohmann, K. K., & Katsos, N. (2014). Is bilectalism similar to bilingualism? An investigation into children's vocabulary and executive control skills. In Will Orman & Matthew James Valleau (Eds.), *BUCLD 38: Proceedings of the 38th Annual Boston University Conference on Language Development* (Vol. 1, pp. 12–24). Cascadilla Press.

Apperly, I. A. (2012). What is "Theory of Mind"? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology*, *65*(5), 825–839. https://doi.org/10.1080/17470218.2012.676055

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential Theory of Mind task. *Cognition*, *106*(3), 1093–1108. https://doi.org/10.1016/j.cognition.2007.05.005

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., & Hyde, J. S. (1993). Processing strategies for time-course data sets in functional mri of the human brain. *Magnetic Resonance in Medicine*, *30*(2), 161–173. http://dx.doi.org/10.1002/mrm.v30:2

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. http://dx.doi.org/10.1016/j.cognition.2010.10.010

Barner, D., Brooks, N., & Bale, A. (2016). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *188*, 87–96. https://doi.org/10.31234/osf.io/jufdb

Barner, D., Hochstein, L. K., Rubenson, M. P., & Bale, A. (2018). Four-year-old children compute scalar implicatures in absence of epistemic reasoning. In K. Syrett & A. Arunachalam (Eds.), *Semantics in acquisition* (pp. 326–349). Benjamins. https://doi.org/10.1075/tilar.24.14bar

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of Theory of Mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813–822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

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

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251. https://doi.org/10.1111/1469-7610.00715

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software*, *67*(1), 48. https://doi.org/10.18637/jss.v067.i01

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

Bitan, T., Burman, D. D., Lu, D., Cone, N. E., Gitelman, D. R., Mesulam, M. M., & Booth, J. R. (2006). Weaker top–down modulation from the left inferior frontal gyrus in children. *Neuroimage*, *33*(3), 991–998. https://doi.org/10.1016/j.neuroimage.2006.07.007

Bonomini, V., Zucchelli, L., Re, R., Ieva, F., Spinelli, L., Contini, D., Paganoni, A., & Torricelli, A. (2015). Linear regression models and k-means clustering for statistical analysis of fNIRS data. *Biomedical Optics Express*, *6*(2), 615. http://dx.doi.org/10.1364/BOE.6.000615

Breheny, R., Ferguson, H., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423–440. https://doi.org/10.1016/j.cognition.2012.11.012

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. (2002). Immature frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron*, *33*(2), 301–311. https://doi.org/10.1016/s0896-6273(01)00583-9

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, *4*(5), e125. https://doi.org/10.1371/journal.pbio.0040125

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595–616. https://doi.org/10.1207/s15326942dn2802_3

Carrion, V. G., Garrett, A., Menon, V., Weems, C. F., & Reiss, A. L. (2008). Posttraumatic stress symptoms and brain function during a response-inhibition task: An fMRI study in youth. *Depression and Anxiety*, *25*(6), 514–526. https://doi.org/10.1002/da.20346

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. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. *Structures and Beyond*, *3*, 39–103. https://doi.org/10.1093/oso/9780195171976.003.0003

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, *37*(4), 535–590. https://doi.org/10.1162/ling.2006.37.4.535

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *BUCLD 25 proceedings* (pp. 157–168). Cascadilla Press.

Chierchia, G., Fox, D., & Spector, B. (2009). Hurford's constraint and the theory of scalar implicatures. In P. Egré & G. Magri (Eds.), *Presuppositions and implicatures – Proceedings of the MIT-Paris workshop* (pp. 47–62). MIT Working Papers in Linguistics (MITWPL).

Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between syntax and semantics. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 3, pp. 2297–2332). New York: Mouton de Gruyter.

Ciesielski, K. T., Lesnik, P. G., Savoy, R. L., Grant, E. P., & Ahlfors, S. P. (2006). Developmental neural networks in children performing a categorical N-Back task. *Neuroimage*, *33*(3), 980–990. https://doi.org/10.1016/j.neuroimage.2006.07.028

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*(*2*), 417–431. http://dx.doi.org/10.1017/S0305000900013842

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Croft, L. J., Baldeweg, T., Sepeta, L., Zimmaro, L., Berl, M. M., & Gaillard, W. D. (2014). Vulnerability of the ventral language network in children with focal epilepsy. *Brain*, *137*(8), 2245–2257. https://doi.org/10.1093/brain/awu154

Crone, E. A., Donohue, S. E., Honomichl, R., Wendelken, C., & Bunge, S. A. (2006a). Brain regions mediating flexible rule use during development. *The Journal of Neuroscience*, *26*(43), 11239–11247. https://doi.org/10.1523/jneurosci.2165-06.2006

Crone, E. A., Wendelken, C., Donohue, S., Van Leijenhorst, L., & Bunge, S. A. (2006b). Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(24), 9315–9320. https://doi.org/10.1073/pnas.0510088103

Crone, E. A., Zanolie, K., Van Leijenhorst, L., Westenberg, M., & Rombouts, S. A. R. B. (2008). Neural mechanisms supporting flexible performance adjustment during development. *Cognitive, Affective, and Behavioral Neuroscience*, *8*(2), 165–177. https://doi.org/10.3758/cabn.8.2.165

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage*, *54*(4), 2808–2821. https://doi.org/10.1016/j.neuroimage.2010.10.069

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system® (D-KEFS®): Examiner's manual: Flexibility of thinking, concept formation, problem solving, planning, creativity, impulse control, inhibition*. Pearson.

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. https://doi.org/10.1027/1618-3169.54.2.128

Dibbets, P., Bakker, K., & Jolles, J. (2006). Functional MRI of task switching in children with specific language impairment (SLI). *Neurocase*, *12*(1), 71–79. https://doi.org/10.1080/13554790500507032

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*(*5*), 630–641. http://dx.doi.org/10.1037//0012-1649.37.5.630

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/bf03203267

Fairchild, S., & Papafragou, A. (2021). The role of executive function and Theory of Mind in pragmatic computations. *Cognitive Science*, *45*(2), e12938. https://doi.org/10.1111/cogs.12938

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyzes using G * Power 3.1: Tests for correlation and regression analyzes. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fishburn, F. A., Norr, M. E., Medvedev, A. V., & Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Frontiers in Human Neuroscience*, *8*, 76. https://doi.org/10.3389/fnhum.2014.00076

Flandin, G., & Novak, M. J. (2020). fMRI data analysis using SPM. In Stephen Ulmer & Olav Jansen (Eds.), *fMRI* (pp. 89–116). Springer. https://doi.org/10.1007/978-3-030-41874-8_8

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

Foppolo, F., Mazzaggio, G., Panzeri, F., & Surian, L. (2021). Scalar and Ad-Hoc pragmatic inferences in children: Guess which one is easier. *Journal of Child Language*, *48*(2), 350–372. https://doi.org/10.1017/s030500092000032x

Fox, D. (2007). Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (Eds.), *Presupposition and implicature in compositional semantics* (pp. 71–120). Palgrave Macmillan. https://doi.org/10.1057/9780230210752_4

Gallagher, A., Béland, R., & Lassonde, M. (2012). The contribution of functional near-infrared spectroscopy (fNIRS) to the presurgical assessment of language function in children. *Brain and Language*, *121*(2), 124–129. https://doi.org/10.1016/j.bandl.2011.03.006

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gernsbacher, M. A., & Yergeau, M. (2019). Empirical failures of the claim that autistic people lack a Theory of Mind. *Archives of Scientific Psychology*, *7*(1), 102–118. https://doi.org/10.1037/arc0000067

Gershon, R. C., Cook, K. F., Mungas, D., Manly, J. J., Slotkin, J., Beaumont, J. L., & Weintraub, S. (2014). Language measures of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society*, *20*(6), 642–651. https://doi.org/10.1017/s1355617714000411

Gleason, J. B. (2005). *The development of language* (6th ed.). Pearson.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (Vol. 3). Academic Press.

Grice, P. (1989). *Studies in the way of words*. Harvard University Press.

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, *65*(6), 852–858. https://doi.org/10.1016/j.neuron.2010.03.001

Grossmann, T., Parise, E., & Friederici, A. (2010). The detection of communicative signals directed at the self in infant prefrontal cortex. *Frontiers in Human Neuroscience*, *4*, 201. https://doi.org/10.3389/fnhum.2010.00201

Guasti, M. T., 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

Gweon, H., Dodell-Feder, D., Bedny, M., & Saxe, R. (2012). Theory of Mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Development*, *83*(6), 1853–1868. https://doi.org/10.1111/j.1467-8624.2012.01829.x

Happé, F. G. (1994). An advanced test of Theory of Mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

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

Heyman, T., & Schaeken, W. (2015). Some differences in some: Examining variability in the interpretation of scalars using latent class analysis. *Psychologica Belgica*, *55*(1), 1–18. https://doi.org/10.5334/pb.bc

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(*1-2*), 67–99. http://dx.doi.org/10.1016/j.cognition.2003.10.011

Hirschberg, J. L. B. (1985). *A theory of scalar implicature* [Unpublished doctoral dissertation]. University of Pennsylvania.

Hochstein, L., Bale, A., & Barner, D. (2018). 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

Hochstein, L., Bale, A., Fox, D., & Barner, D. (2014). Ignorance and inference: Do problems with Gricean epistemic reasoning explain children's difficulty with scalar implicature? *Journal of Semantics*, *33*, 107–135. https://doi.org/10.1093/jos/ffu015

Hollebrandse, B., van Hout, A., & Hendriks, P. (2014). Children's first and second-order false-belief reasoning in a verbal and a low-verbal task. *Synthese*, *191*(3), 321–333. https://doi.org/10.1007/s11229-012-0169-9

Horn, L. (1989). *A natural history of negation*. University of Chicago Press.

Houdé, O., Rossi, S., Lubin, A., & Joliot, M. (2010). Mapping numerical processing, reading, and executive functions in the developing brain: An fMRI meta-analysis of 52 studies including 842 children. *Developmental Science*, *13*(6), 876–885. https://doi.org/10.1111/j.1467-7687.2009.00938.x

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

Huppert, T., Barker, J., Schmidt, B., Walls, S., & Ghuman, A. (2017). Comparison of group-level, source localized activity for simultaneous functional near-infrared spectroscopy-magnetoencephalography and simultaneous fNIRS-fMRI during parametric median nerve stimulation. *Neurophotonics*, *4*(1), 015001–015001. https://doi.org/10.1117/1.NPh.4.1.015001

Issard, C., & Gervain, J. (2017). Adult-like processing of time-compressed speech by newborns: A NIRS study. *Developmental Cognitive Neuroscience*, *25*, 176–184. https://doi.org/10.1016/j.dcn.2016.10.006

Issard, C., & Gervain, J. (2018). Variability of the hemodynamic response in infants: Influence of experimental design and stimulus complexity. *Developmental Cognitive Neuroscience*, *33*, 182–193. https://doi.org/10.1016/j.dcn.2018.01.009

Jang, K. E., Tak, S., Jung, J., Jang, J., Jeong, Y., & Ye, J. C. (2009). Wavelet-MDL detrending for near infrared spectroscopy (NIRS). *Journal of Biomedical Optics*, *14*(3), 1–13. https://doi.org/10.1117/12.764141

Jasińska, K. K., Berens, M. S., Kovelman, I., & Petitto, L. A. (2017). Bilingualism yields language-specific plasticity in left hemisphere's circuitry for learning to read in young children. *Neuropsychologia*, *98*, 34–45. https://doi.org/10.1016/j.neuropsychologia.2016.11.018

Jasińska, K. K., & Petitto, L. A. (2013). How age of bilingual exposure can change the neural systems for language in the developing brain: A functional near infrared spectroscopy investigation of syntactic processing in monolingual and bilingual children. *Developmental Cognitive Neuroscience*, *6*, 87–101. https://doi.org/10.1016/j.dcn.2013.06.005

Jasińska, K. K., & Petitto, L. A. (2014). Development of neural systems for reading in the monolingual and bilingual brain: New insights from functional near infrared spectroscopy neuroimaging. *Developmental Neuropsychology*, *39*(6), 421–439. https://doi.org/10.1080/87565641.2014.939180

Jasińska, K. K., Shuai, L., Lau, A. N., Frost, S., Landi, N., & Pugh, K. R. (2021). Functional connectivity in the developing language network in 4-year-old children predicts future reading ability. *Developmental Science*, *24*(2), e13041. https://doi.org/10.1111/desc.13041

Jasper, H. H. (1961). The ten twenty electrode system: International federation of societies for electroencephalography and clinical neurophysiology. *American Journal of EEG Technology*, *1*(2), 13–19. https://doi.org/10.1080/00029238.1961.11080571

Kampa, A., & Papafragou, A. (2020). Four-year-olds incorporate speaker knowledge into pragmatic inferences. *Developmental Science*, *23*(3), e12920. https://doi.org/10.1111/desc.12920

Kampa, A., & Papafragou, A. (2023). Children and adults use pragmatic principles to interpret non-linguistic symbols. *Journal of Memory and Language*, *131*(5), 104429. https://doi.org/10.1016/j.jml.2023.104429

Kampa, A., Richards, C., & Papafragou, A. (2024). Preschool children generate quantity inferences from both words and pictures. *Journal of Experimental Child Psychology*, *239*, 105805. https://doi.org/10.1016/j.jecp.2023.105805

Katsos, N., & Bishop, D. V. (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

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kersey, A., & Emberson, L. (2017). Tracing trajectories of audio-visual learning in the infant brain. *Developmental Science*, *20*(6), e12480. https://doi.org/doi:10.1111/desc.12480

Kohl, S. H., Mehler, D. M. A., Lührs, M., Thibault, R. T., Konrad, K., & Sorger, B. (2020). The potential of functional near-infrared spectroscopy-based neurofeedback—A systematic review and recommendations for best practice. *Frontiers in Neuroscience*, *14*, 594. https://doi.org/10.3389/fnins.2020.00594

Levinson, S. (2000). *Presumptive meanings*. MIT Press. https://doi.org/10.7551/mitpress/5526.001.0001

Marty, P. P., & Chemla, E. (2013). Scalar implicatures: Working memory and a comparison with only. *Frontiers in Psychology*, *4*, 403. https://doi.org/10.3389/fpsyg.2013.00403

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

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, *27*(*3*), 403–422. http://dx.doi.org/10.1017/S0142716406060334

Molfese, P., Glen, D., Mesite, L., Cox, R., Hoeft, F., Frost, S., Mencl, W. E., Pugh, K., & Bandettini, P. (2021). The Haskins pediatric atlas: a magnetic-resonance-imaging-based pediatric template and atlas. *Pediatric Radiology*, *51*(4), 628–639. https://doi.org/10.1007/s00247-020-04958-w

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(*4*), 329–336. http://dx.doi.org/10.1111/j.0956-7976.2002.00460.x

Nichols, T. (2013). *False discovery rate*. Retrieved February, 2021, from http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/software/fdr

Nishiyori, R. (2016). fNIRS: An emergent method to document functional cortical activity during infant movements. *Frontiers in Psychology*, *7*, 533. https://doi.org/10.3389/fpsyg.2016.00533

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

Noveck, I. A., Guelminger, R., Georgieff, N., & Labruyere, N. (2007). What autism can reveal about every … not sentences. *Journal of Semantics*, *24*(1), 73–90. https://doi.org/10.1093/jos/ffl009

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203–210. https://doi.org/10.1016/s0093-934x(03)00053-1

O'Hare, E. D., Lu, L. H., Houston, S. M., Bookheimer, S. Y., & Sowell, E. R. (2008). Neurodevelopmental changes in verbal working memory load-dependency: An fMRI investigation. *NeuroImage*, *42*(4), 1678–1685. https://doi.org/10.1016/j.neuroimage.2008.05.057

Olesen, P. J., Macoveanu, J., Tegner, J., & Klingberg, T. (2006). Brain activity related to working memory and distraction in children and adults. *Cerebral Cortex*, *17*(5), 1047–1054. https://doi.org/10.1093/cercor/bhl014

O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*(*2*), 659. http://dx.doi.org/10.2307/1131839

Papafragou, A., Friedberg, C., & Cohen, M. (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

Pereira, J., Direito, B., Lührs, M., Castelo-Branco, M., & Sousa, T. (2023). Multimodal assessment of the spatial correspondence between fNIRS and fMRI hemodynamic responses in motor tasks. *Scientific Reports*, *13*(1), 2244. https://doi.org/10.1038/s41598-023-29123-9

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

Petroni, A., Cohen, S. S., Ai, L., Langer, N., Henin, S., Vanderwal, T., Milham, M. P., & Parra, L. C. (2018). The variability of neural responses to naturalistic videos change with age and sex. *eNeuro*, *5*(*1*), ENEURO.0244-17.2017. http://dx.doi.org/10.1523/ENEURO.0244-17.2017

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. https://doi.org/10.1007/s10803-008-0661-8

Politzer-Ahles, S., Fiorentino, R., Jiang, X., & Zhou, X. (2013). Distinct neural correlates for pragmatic and semantic meaning processing: An event-related potential investigation of scalar implicature processing using picture-sentence verification. *Brain Research*, *1490*, 134–152. https://doi.org/10.1016/j.brainres.2012.10.042

Politzer-Ahles, S., & Gwilliams, L. (2015). Involvement of prefrontal cortex in scalar implicatures: Evidence from magnetoencephalography. *Language, Cognition and Neuroscience*, *30*(7), 853–866. https://doi.org/10.1080/23273798.2015.1027235

Pouscoulous, N., Noveck, I. A., 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

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Quaresima, V., Bisconti, S., & Ferrari, M. (2012). A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain and Language*, *121*(2), 79–89. https://doi.org/10.1016/j.bandl.2011.03.009

Rollins, P. R. (1999). Early pragmatic accomplishments and vocabulary development in preschool children with autism. *American Journal of Speech-Language Pathology*, *8*(2), 181–190. https://doi.org/10.1044/1058-0360.0802.181

Sassa, Y., Sugiura, M., Jeong, H., Horie, K., Sato, S., & Kawashima, R. (2007). Cortical mechanism of communicative speech production. *Neuroimage*, *37*(3), 985–992. https://doi.org/10.1016/j.neuroimage.2007.05.059

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, *27*(3), 367–391. https://doi.org/10.1023/b:ling.0000023378.71748.db

Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for Theory of Mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience*, *1*(3), 229–234. https://doi.org/10.1093/scan/nsl034

Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K. A. (2009). Brain regions for perceiving and reasoning about other people in school-aged children. *Child Development*, *80*(4), 1197–1209. https://doi.org/10.1111/j.1467-8624.2009.01325.x

Schulz, K. P., Fan, J., Tang, C. Y., Newcorn, J. H., Buchsbaum, M. S., Cheung, A. M., & Halperin, J. M. (2004). Response inhibition in adolescents diagnosed with attention deficit hyperactivity disorder during childhood: An event-related FMRI study. *American Journal of Psychiatry*, *161*(9), 1650–1657. https://doi.org/10.1176/appi.ajp.161.9.1650

Shalinsky, M. H., Kovelman, I., Berens, M. S., & Petitto, L. A. (2009). Exploring cognitive functions in babies, children & adults with near infrared spectroscopy. *Journal of Visualized Experiments*, *29*, e1268. https://doi.org/10.3791/1268

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

Shetreet, E., Chierchia, G., & Gaab, N. (2014b). Linguistic inability or poor performance: Dissociating scalar implicature generation and mismatch in the developing brain. *Developmental Psychology*, *50*(9), 2264–2275. https://doi.org/10.1037/a0037368

Singh, A. K., Okamoto, M., Dan, H., Jurcak, V., & Dan, I. (2005). Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. *Neuroimage*, *27*(4), 842–851. https://doi.org/10.1016/j.neuroimage.2005.05.019

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. (1995). *Relevance: Communication and cognition* (2nd ed.). Harvard University Press. (Original work published 1986).

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, *17*(1-2), 3–23. https://doi.org/10.1111/1468-0017.00186

Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgment task: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, *31*(6), 817–840. https://doi.org/10.1080/23273798.2016.1161806

Spychalska, M., Reimer, L., Schumacher, P. B., & Werning, M. (2021). The cost of the epistemic step: Investigating scalar implicatures in full and partial information contexts. *Frontiers in Psychology*, *12*, 2133. https://doi.org/10.3389/fpsyg.2021.679491

Stiller, A., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*(2), 176–190. https://doi.org/10.1080/15475441.2014.927328

Taga, G., Asakawa, K., Hirasawa, K., & Konishi, Y. (2003). Hemodynamic responses to visual stimulation in occipital and frontal cortex of newborn infants: A near-infrared optical topography study. *Early Human Development*, *75*, 203–210. https://doi.org/10.1016/j.earlhumdev.2003.08.023

Taga, G., Asakawa, K., Maki, A., Konishi, Y., & Koizumi, H. (2003). Brain imaging in awake infants by near-infrared optical topography. *Proceedings of the National Academy of Sciences*, *100*(19), 10722–10727. https://doi.org/10.1073/pnas.1932552100

Tian, X., Hao, X., Song, Y., & Liu, J. (2021). Homogenization of face neural representation during development. *Developmental Cognitive Neuroscience*, *52*, 101040. https://doi.org/10.1016/j.dcn.2021.101040

Van Tiel, B., & Kissine, M. (2018). Quantity-based reasoning in the broader autism phenotype: A web-based study. *Applied Psycholinguistics*, *39*(6), 1373–1403. https://doi.org/10.1017/s014271641800036x

Wagstaff, K., Cardie, C., Rogers, S., & Schrödl, S. (2001, June). Constrained k-means clustering with background knowledge. *ICML*, *1*, 577–584.

Walsh, B., Tian, F., Tourville, J. A., Yücel, M. A., Kuczek, T., & Bostian, A. J. (2017). Hemodynamics of speech production: An fNIRS investigation of children who stutter. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-04357-6

Watanabe, H., Homae, F., & Taga, G. (2012). Activation and deactivation in response to visual stimulation in the occipital cortex of 6-month-old human infants. *Developmental Psychobiology*, *54*(1), 1–15. https://doi.org/10.1002/dev.20569

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., … Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(11 Suppl. 3), S54–S64.

Weiss-Croft, L. J., & Baldeweg, T. (2015). Maturation of language networks in children: A systematic review of 22 years of functional MRI. *NeuroImage*, *123*, 269–281. https://doi.org/10.1016/j.neuroimage.2015.07.046

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

Wild, C. J., Linke, A. C., Zubiaurre-Elorza, L., Herzmann, C., Duffy, H., Han, V. K., Lee, D. S. C., & Cusack, R. (2017). Adult-like processing of naturalistic sounds in auditory cortex by 3-and 9-month old infants. *NeuroImage*, *157*, 623–634. https://doi.org/10.1016/j.neuroimage.2017.06.038

Wilke, M., Lidzba, K., & Krägeloh-Mann, I. (2009). Combined functional and causal connectivity analyses of language networks in children: A feasibility study. *Brain and Language*, *108*(1), 22–29. https://doi.org/10.1016/j.bandl.2008.09.007

Willems, R. M., & Varley, R. A. (2010). Neural insights into the relation between language and communication. *Frontiers in Human Neuroscience*, *4*, 203. https://doi.org/10.3389/fnhum.2010.00203

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited—again. *NeuroImage*, *2*(*3*), 173–181. http://dx.doi.org/10.1006/nimg.1995.1023

Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. *Neuroimage*, *44*(2), 428–447. https://doi.org/10.1016/j.neuroimage.2008.08.036

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., Conway, K. P., Gershon, R., & Weintraub, S. (2014). NIH toolbox cognition battery (CB): Validation of executive function measures in adults. *Journal of the International Neuropsychological Society*, *20*(6), 620–629. https://doi.org/10.1017/s1355617714000472

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37–63. https://doi.org/10.1016/s0885-2014(96)90027-1

Zelazo, P. D., & Müller, U. (2010). Executive function in typical and atypical development. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 574–603). Wiley-Blackwell. https://doi.org/10.1002/9781444325485.ch22