The neural development of pragmatic inference-making
in natural discourse

Flora Schwartz, Justine Epinat-Duclos, Ira Noveck, Jérôme Prado

HAL Id: hal-02346617
https://hal.science/hal-02346617

Submitted on 5 Nov 2019

The neural development of pragmatic inference-making in natural discourse

Flora Schwartz[1], Justine Epinat-Duclos[1], Ira Noveck[1], & Jérôme Prado[1]

[1] Institut des Sciences Cognitives Marc Jeannerod – UMR 5304, Centre National de la Recherche Scientifique (CNRS) & Université de Lyon, France

Author Note

Correspondence concerning this article should be addressed to Flora Schwartz (flora.schwartz@isc.cnrs.fr) or Jérôme Prado (jprado@isc.cnrs.fr), Institut des Sciences Cognitives Marc Jeannerod – UMR 5304, Centre National de la Recherche Scientifique (CNRS) & Université de Lyon, 67 Boulevard Pinel, 69675 Bron cedex, France.

Word count: 7922 words

Abstract

Older interlocutors are more likely than younger ones to make pragmatic inferences, i.e., inferences that go beyond the linguistically encoded meaning of a sentence. Here we ask whether pragmatic development is associated with increased activity in brain structures associated with inference-making or in those associated with Theory-of-Mind. We employed a reading task that presents vignettes in one of two versions, one of which is expected to prompt more pragmatic processing. Both versions present a major premise containing three possibilities (e.g. Xavier is considering Thursday, Friday or Saturday for inviting his girlfriend out). In the *Fully-Deductive* (control) condition, the major premise is followed by two disjunction-elimination premises through two separate lines (one indicating that Saturday is not convenient and another saying that Thursday is not convenient); this is followed by a valid conclusion (e.g., "I'll reserve Friday"). In the *Implicated-Premise* condition, the first disjunction-elimination premise is followed by a second similar one that eliminates the same disjunction (e.g. both lines explain why Saturday is not convenient). In this condition, readers may pragmatically enrich the conclusion (i.e. "I'll reserve Friday" pragmatically implicates that *Xavier is also eliminating Thursday from consideration*). Reading times in Experiment 1 showed that processing the speaker's conclusion in the *Implicated-Premise* condition becomes increasingly more effort-demanding as readers reach adolescence. Experiment 2 showed that this developmental pattern is related to age-related increases of fMRI activity in fronto-parietal regions typically involved in inference-making processes. We found no evidence indicating age effects related to Theory of Mind areas.

Introduction

Pragmatics is concerned with the way a listener goes beyond what a speaker *says* in order to understand what the speaker *means* (Grice, 1989). It covers a wide range of readily perceptible phenomena, such as metaphor (*John is a cheetah* to describe a fast runner) and irony (*that went well* to describe a failed meeting) as well as more subtle cases such as commands (consider how *Pick up the tall glass* can indicate to the listener that there is a smaller glass in the speaker's frame of reference). Common to all these cases is the fact that the linguistically encoded meaning of the utterance serves as a basis for inferring what the speaker intended.

Developmental progressions have been crucial to investigating pragmatics. The most discussed example in the experimental pragmatic literature focuses on underinformative utterances such as *Some swans are white*, which can be enriched to mean *Some but not all swans are white* (for reviews, see Noveck & Reboul, 2008; Papafragou, 2015). Consider the work from Pouscoulous et al. (2007) who showed 9-year-old and adult participants toy animals that were placed in and around a set of four boxes; one observable fact was that there was a turtle in each of the boxes and nowhere else. When participants were asked to evaluate the sentence "Some turtles are in the boxes" as true or false, nine-year-olds were significantly more likely than adults to answer true. That is because children appear satisfied with the semantic reading of *Some* (*Some and perhaps all*) while adults are more likely to pragmatically enrich the utterance to *Some but not all boxes have a turtle* (also see Huang & Snedeker, 2009; Katsos et al., 2016). That pragmatic abilities become increasingly evident with age can be extended to a range of phenomena, including irony (Fillipova & Astington, 2010), metaphoric reference (Noveck et al., 2001; Almor et al., 2007) and idioms (Caillies & Le Sourn-Bissaoui, 2008).

Given the plethora of studies that reveal a progressive emergence of pragmatic abilities among children, the question we ask here is a general one: what are the neural structures that subserve this development? We consider two possible neural systems in order to answer this question. One is derived by assuming that developmental changes arise due to generally improving inferential abilities as indicated by neuroimaging studies that point out how lateral frontoparietal regions known to support inference-making in adults (Monti et al., 2007; 2009; Prado et al., 2011; Reverberi et al., 2010; Vendetti & Bunge, 2014) *mature* from middle childhood to adolescence. For example, studies on analogical reasoning have demonstrated relatively late age differences in the bilateral rostrolateral prefrontal cortex (RLPFC), dorsolateral prefrontal cortex (DLPFC), and inferior parietal lobule (IPL) (Crone et al., 2009; Wendelken et al., 2011, Dumontheil et al., 2010), areas that are thought to support relational integration (Dumontheil, 2014) and working-memory demands (Badre & D'Esposito, 2007) of inference-making. Age-related increases of activity between middle childhood and adolescence have also been found in the RLPFC and DLPFC when participants make deductive inferences such as *All As are Bs* and *All Bs are Cs* therefore *All As are Cs* (Mathieu et al., 2015). Finally, anatomical studies show structural changes (e.g., an increase in white matter density) in the lateral prefrontal cortex until adolescence (Barnea-Goraly et al., 2005; Westlye et al., 2010). Overall, neuroimaging studies provide evidence for a relatively late functional and anatomical maturation of the lateral frontoparietal regions subserving inference-making. The maturation of this *inference-making network* may be a critical factor underlying the development of pragmatic skills in children and adolescents[1].

Another possibility is that emerging pragmatic abilities are linked to a developing Theory of Mind (ToM) (Epley et al., 2004b; Dumontheil et al., 2010) because a critical

---

[1] Even though we refer here to this set of lateral frontoparietal regions as an "inference-making network", it should be emphasized that we do not believe this network is specifically involved in inference-making. Rather, the network is more likely involved in domain-general processes *involved* in inference-making.

feature of understanding a speaker's intended meaning is the attribution of mental states to others. Although ToM-related skills, such as perspective taking (understanding other's objectives and preferences) and the evaluation of false beliefs (i.e. distinguishing between what someone believes and reality), appear to develop relatively early in children (Wellman et al., 2001; Baillergeon et al., 2010; Southgate et al., 2007), there are also studies suggesting that ToM continues to develop well into adolescence (Dumontheil et al., 2010). For example, Dumontheil et al (2010) showed a late developmental improvement in the ability to take into account a speaker's perspective in an online communication game (Keysar et al., 2000; 2003). This late development is also consistent with two neuroimaging studies indicating that several of the brain regions thought to be involved in ToM are not mature until late childhood (Mills et al., 2014; Saxe et al., 2009). For example, Saxe et al. (2009) found functional changes between 6 and 11 years of age in the temporo-parietal junction (TPJ), a core region supporting ToM (Van Overwalle, 2009). Additionally, Mills et al (2014) showed an increase in gray matter volume among children between 7 and 10 years of age in the medial prefrontal cortex (mPFC), posterior superior temporal sulcus (pSTS), and TPJ, three core regions of ToM (Van Overwalle, 2009). As far as pragmatic text comprehension tasks go, it has been demonstrated how brain regions associated with Theory of Mind are central to irony understanding in adults (Spotorno et al., 2012); given that the development of irony has been documented, it would make sense to suppose that children's weaknesses with respect to irony are due to a lack of maturity of ToM. It is thus very reasonable to hypothesize that the occurrence of pragmatic competence with age points specifically to maturing ToM regions.

Here we aim to test between these two hypotheses by identifying the brain regions supporting the development of pragmatic inference-making in children and adolescents. We adopted a text-comprehension task that our group has used recently with adults for the neural investigation of pragmatic inference-making (Prado et al., 2015). This paradigm presents a

series of stories that include an utterance stated by a story character (1f below) that could be interpreted in one of two ways as a function of a small change in the preceding context. The first sort of interpretation is referred to as *Fully-Deductive* because the nature of a character's remark is, in effect, a conclusion drawn from two successive disjunction-elimination inferences, as exemplified in lines (1c) through (1e), that reduce three possibilities to one:

(1)   (a) Xavier wants to surprise Claire by inviting her to a restaurant.

(b) He hesitates between these three days of the week:

(c) Thursday, Friday or Saturday.

(d) Claire cannot go on Saturday because she has a theatre course.

(e) Further, Thursday would not work because he is having dinner with his parents.

(f) He says: "I'll invite her for Friday."

(g) He hopes she will be available that day.

As this story (translated from French) exemplifies, it begins with an introduction setting up a context (in lines 1a and 1b) before presenting a disjunctive statement in which three possibilities are mentioned (in 1c). Two of these possibilities are then eliminated by the speaker (in 1d and 1e) such that its conclusion (in 1f), which logically follows, serves as a kind of verification.

The other sort of interpretation is referred to as an *Implicated-Premise* interpretation. There, the story sets up the context in such a way that the line in (1f) is not conclusive; that is, only one possibility of three is eliminated. To arrive at this sort of interpretation, consider the above story in (1), but replace the word "Thursday" with "Saturday" in line (1e) so that it now reads as:

(1)   (e') Further, Saturday would not work because he is having dinner with his parents.

In other words, *Implicated-Premise* stories provide two justifications for eliminating just one possibility before presenting the story-character's utterance. Now, Xavier's conclusion in (1f) -- that Friday is a good day for the invitation -- still makes sense, but can be understood in one of two ways: A participant could process (1f) without a deep reading, i.e. by simply accepting the speaker's utterance without giving further thought to the speaker's reasoning. In this way, the linguistically encoded reading is in effect good enough for processing its meaning. The other way to process (1f) is to note that the conclusion *puts the reader in a position to infer* that Xavier is eliminating Thursday, or must have a reason to eliminate Thursday, even though he never said so explicitly. This is a pragmatic inference on the part of the reader, whose effort to understand the utterance is part of an attempt to access the speaker's intention. Note that this is not an obligatory inference on a par with a deductive one. Our prior study (Prado et al, 2015) showed that adults take longer to read the critical line (1f) when it is a potential source for an *Implicated premise* (in the context of a line like [1e']) than they do when it is part of a *Fully-Deductive* conclusion (in the context of a line like [1e]).

In turning to our current developmental study, the goal is twofold. First, in line with other developmental pragmatic phenomena, we determine whether the reading times of critical utterances (like those in line [1f]) in the *Implicated-Premise* versions will become progressively more distinctive from those in the *Fully-Deductive* versions as participants get older; in other words, lines like (1f) in the *Implicated-Premise* condition will more likely be read as an unremarkable assertion (i.e. without pragmatic enrichment) among younger children and as a source of further pragmatic enrichment among older children (and adolescents). In Experiment 1, we first establish this anticipated result through a behavioral experiment in which children (ages 8 to 15) and adults read *Fully-Deductive* and *Implicated-*

6

Premise stories while we track the latencies of the critical utterances in the sixth line of each story. In our critical follow-up experiment, we use the same paradigm and fMRI in children from 8 to 15 while investigating age-related changes of activity in brain regions supporting general inference-making (the RLPFC, DLPFC and IPL) and those supporting ToM (the bilateral TPJ, the mPFC and the PC).

Experiment 1

Material and Methods

Participants

Sixty-one healthy adults (age range: 21-45 years) and 91 typically developing children participated in the experiment. There were 19 third-graders (age range: 8-9 years), 30 fifth-graders (age range: 10-11 years), 20 seventh-graders (age range: 12-13 years), and 22 ninth-graders (age range: 14-15 years). All participants were native French speakers. Nineteen participants were excluded from further analyses because of performance below or equal to 50% on the comprehension questions. An additional participant was excluded because of very slow (>10s) or very fast (<1s) reading times on 5 out of 12 of the critical sentences. The final dataset included 15 third-graders, 25 fifth-graders, 18 seventh-graders, 18 ninth-graders, and 56 adults.

Materials

Participants were presented with two sorts of short stories containing a disjunction-elimination argument similar to the scenarios in (1) as described in the Introduction (see **Table 1**). All stories were composed of 7 sentences each, with a mean length of 376 characters per sentence (standard deviation [sd]=28.5), and had the same structure: Sentence (a) introduced the context, sentence (b) introduced the disjunctive argument, and sentences (c) through (e) represented the main premises of the argument. Specifically, sentence (b) always pointed out that there were three possibilities that could account for an event (e.g., going out

7

on one of three days of the week). Sentence (c) enumerated these possibilities (e.g., "Thursday, Friday or Saturday") and sentence (d) ruled out one of these possibilities (e.g., "not on Saturday"). The key manipulation was in sentence (e). In half of the stories (i.e., *Fully-Deductive* stories), sentence (e) ruled out a second possibility out of the three (e.g., "not on Thursday"). In the other half of the stories (i.e., *Implicated-Premise* stories), sentence (e) gave another reason for eliminating the possibility that had been already ruled out in sentence (d) (i.e., confirming "not on Saturday"). Sentence (f) presented the story character's critical remark (e.g., "it will be on Friday"). Therefore, in *Fully-Deductive* stories, the utterance necessarily follows from the premises and is confirmatory. In *Implicated-Premise* stories, participants read the protagonist's conclusion and are in a position to draw a further inference that would justify the conclusion. Sentence (g) was the wrap-up sentence. In general, only one word differed between *Fully-Deductive* and *Implicated-Premise* stories (in sentence (e)). Seventeen different story frameworks were created. Each framework involved a different situation and different characters. Each participant read either the *Fully-Deductive* or the *Implicated-Premise* version of a framework.

Twelve stories were randomly chosen for each participant and these were evenly divided between *Fully-Deductive* and *Implicated-Premise* stories. Two-thirds of *Fully-Deductive* and *Implicated-Premise* stories were followed by a yes/no comprehension question (e.g., "In your opinion, does Xavier plan to invite his girlfriend?", see **Table 1**). This question was designed to verify that participants were paying attention to the content in the stories. One-third of *Fully-Deductive* and *Implicated-Premise* stories were followed by a question that required participants to judge whether the conclusion made by the character was justified (e.g., "In your opinion, is Xavier justified to think that he should invite Claire on Friday?", see **Table 1**). The justification question provides an off-line measure to determine whether participants detect the difference between the two types of story.

Experimental procedures

The experiment was programmed with MATLAB while using the Psychophysics Toolbox (Brainard, 1997). Stories were presented line by line on a computer screen. Participants were instructed to read at their own pace and to press the space bar to read the next sentence. Each story started with a red fixation cross, which appeared on the left side of the screen. The cross remained red for 3 seconds, orange for 1 second and green for 1 second. Each line of the story was then displayed in a left-justified manner half-way up the screen. Each sentence remained on the screen until the participant pressed a key, or after 12 seconds if no key was pressed. A black screen was displayed for a half-second between each sentence. After the disappearance of the last line, a white fixation cross appeared on the left of the screen for 2 seconds. The question then appeared at the center of the screen. Participants had a 12 second window to answer the question. The trial ended with the participant's response or after the allotted time if no response was provided. Each story was directly followed by a period of visual fixation (ranging from 2 to 4 seconds). All stories were presented in a randomized order.

Analyses

Statistical analyses were conducted using the lme4 package implemented in R (Bates, Maechler & Bolker, 2012). Log-transformed reading times were analyzed using a linear mixed model, while responses to comprehension and justification questions were analyzed using logistic mixed models. All full models included in their fixed effects an intercept, a main effect of Story version (i.e., *Fully-Deductive* versus *Implicated-Premise*), a main effect of Age, and an interaction between Story version and Age. The categorical factor Story version was deviation coded as -0.5 for *Fully-Deductive* stories and 0.5 for *Implicated-Premise* stories. Age was entered as a continuous variable by coding third-graders, fifth-graders, seventh-graders, ninth-graders and adults as 8.5, 10.5, 12.5, 14.5 and 25 years

(respectively)[2]. The continuous variable Age was then mean-centered. All models also had maximal random effects structure, with by-subject random intercept and slope for Story version, as well as by-story random intercepts and slopes for Story version, Age, and their interaction. Main effects and interactions were tested using likelihood ratio tests between mixed effect models differing only in the presence or absence of fixed effects of interest. The dataset and R scripts are available at Mendeley Data: http://dx.doi.org/10.17632/fd7t5h3grb.2.

Results and discussion

The linear mixed model with reading time of the conclusion (i.e., sentence (f)) as dependent variable revealed a significant main effect of Age (β=-0.02, SE=0.004, $\chi^2(1)$=21.18, p< 0.001), showing that overall reading time of the conclusion decreased with age. The main effect of Story version was significant (β=0.06, SE=0.03, $\chi^2(1)$=3.93, p=0.047). More importantly, there was a significant interaction between Age and Story version (β=0.012, SE=0.005, $\chi^2(1)$=5.96, p=0.015). That is, the difference in reading time between conclusions of *Fully-Deductive* and *Implicated-Premise* stories increased with age (see **Fig. 1A**). We then tested for the presence of an effect of Story version in each age group separately. Follow-up linear mixed models with the fixed effect Story version (as well as by-subject and by-story random intercepts and slopes for Story version) revealed a significant main effect of Story version (β=0.14, SE=0.03, $\chi^2(1)$=13.82, p< 0.001) for adult participants, but not for any other age groups (all $\chi^2(1)$s < 2.68, all ps> 0.10). In other words, only adults read conclusions of *Implicated-Premise* stories significantly slower than conclusions of *Fully-Deductive* stories.

Overall mean accuracy for comprehension questions was 74% for 3rd graders, 79% for 5th graders, 78% for 7th graders, 88% for 9th graders and 86% for adults. The logistic mixed

---

[2] Note that for illustration purpose descriptive analyses are presented by age-group in all figures.

model with accuracy for comprehension questions as dependent variable revealed a significant main effect of Age (β=0.040, SE=0.013, χ^2(1)=9.79, p=0.002), indicating that overall accuracy increased with age. Neither the main effect of Story version (β=-0.23, SE=0.21, χ^2(1)=1.74, p=0.19) nor the interaction between Age and Story version (β=-0.008, SE=0.026, χ^2(1)=0.10, p=0.75) were significant.

Responses to justification questions were then analyzed to determine whether participants' agreement with the conclusion depended upon Story version and Age (see **Fig. 1B**). The logistic mixed model with accuracy for justification questions as dependent variable revealed significant main effects of Age (β=-0.05, SE=0.02, χ^2(1)=5.43, p=0.02) and Story version (β=-1.45, SE=0.34, χ^2(1)=13.50, p <0.001). That is, acceptance rates of justification questions decreased with age and were overall lower for *Implicit-premise* than *Fully-deductive* stories. Critically, the interaction between Age and Story version was also significant (β=-0.17, SE=0.04, χ^2(1)=16.13, p <0.001). That is, there was a greater age-related decrease in acceptance rates of justification questions for *Implicit-premise* than *Fully-deductive* stories. Follow-up logistic mixed models with the fixed effect Story version (as well as by-subject and by-story random intercepts and slopes for Story version) revealed that the main effect of Story version was significant in adults (β=-3.07, SE=0.67, χ^2(1)=23.86, p< 0.001) and in ninth-graders (β=-2.73, SE=1.13, χ^2(1)=11.86, p <0.001) [3], but not in the other groups (all χ^2(1)s <0.93, all ps> 0.33). Therefore, acceptance rates of justification questions were lower for *Implicated-Premise* than *Fully-Deductive* stories in ninth-graders and adults.

The adults' data here are consistent with those from our previous study (Prado et al. 2015), where we found longer reading times for the conclusions in *Implicated-Premise* stories as well as lower acceptance rates for the justification questions when compared to the *Fully-Deductive* stories. Here we extend these results by reporting a developmental divergence in a)

---

[3] For 9th graders, the random slope structure was simplified by removing the by-story random slope for Story version because the full model did not converge.

conclusion-reading times that become significant among adults and in; b) accepting the justification question that begins to emerge at around 9th grade. Importantly, this developmental trend emerged without negatively affecting the participants' overall comprehension of stories (which increased with age).

We can think of at least 3 possible explanations for this developmental trend, which we now consider one at a time. First, the effect might result from the detection of an infelicity earlier in the story. That is, older participants (as compared to younger ones) may find the 5th sentence (sentence (e)) of *Implicated-Premise* stories infelicitous, because it might repeat some information that was already mentioned in the 4th sentence. This might have downstream consequences on reading times of the conclusion and explain why there is increasing divergence between *Implicated-Premise* conclusion-reading times and *Fully-Deductive* conclusion reading-times with age. It is important to note that all of our stories were written with the utmost care so as to minimize such infelicity in sentence (e). That is, the information that was repeated in sentence (e) was always accompanied by further justification. For example, consider these consecutive statements (lines (c) through (e), translated from French) from our "what night to go out story", "what form of transportation shall he use story" and the "what language are they speaking story":

(2)   (c) Thursday, Friday or Saturday.

         (d) Claire cannot (go out) on Saturday because she has a theatre course.

         (e) Besides, Saturday is not convenient because she is going out with her parents.

(3)   (c) By train, by motorcycle, or by car.

         (d) His motorcycle isn't working because it broke down yesterday.

         (e) Besides, he never uses his motorcycle to go on vacation.

(4)   (c) Italian, Portuguese, or Spanish.

(d) It's not Portuguese because my Mom speaks Portuguese, Camille declares.

(e) She adds after: "I am sure it is not Portuguese."

As one can see, each story presents participants with two reasons to eliminate one option. This arguably occurs in everyday life and is not infelicitous *per se*. Nonetheless, we acknowledge that this repetition might contribute to the difference observed between *Implicated-Premise* and *Fully-Deductive* stories.

A second explanation for the observed effect is that, in a way similar to cloze probability effects (Kutas & Hillyard, 1984; Coulson & Van Petten, 2002), premises in *Fully-Deductive* stories may allow sophisticated readers to know which alternative is going to be mentioned in the conclusion. In contrast, premises in *Implicated-Premise* stories may not allow such a prediction. In other words, the most sophisticated readers may be surprised when reading the conclusion in the *Implicated-premise* condition because they are not expecting it. We acknowledge that surprise may *come into play* with age and that it becomes part of the process in arriving at an alternative reading. Note that this explanation would put the effect squarely in the realm of pragmatics and not as an example of sentence-level processing. To put it in Gricean terms (1989), one could argue that our more sophisticated participants recognize that a maxim (of Relevance) has been violated and that this generates further implicatures in order to maintain the Cooperative Principle. Regardless, a participant would be put in the position to consider what happened to the remaining possibility. The question is whether readers are more likely to pursue that consideration with age.

A third explanation, which we proposed in our previous adult study (Prado et al. 2015), is that conclusions in *Implicated-Premise* stories are more likely than those in the

Fully-Deductive condition to encourage further pragmatic inference-making and particularly among more sophisticated participants. That is, sophisticated readers may make an effort to understand the utterance as part of an attempt to access the speaker's intention. The present experiment is therefore consistent with the idea that pragmatic inference-making emerges over the course of development (Huang & Snedeker, 2009; Katsos et al., 2016; Pouscoulous et al., 2007). If correct, this explanation would raise an intriguing question: which abilities would be responsible for the emergence of such pragmatic inference-making in this task? On the one hand, it may be attributed to the development of inference-making mechanisms over the course of childhood and adolescence. On the other hand, increasingly sophisticated pragmatic abilities may depend upon a more acute ability to attribute mental states to others (i.e., ToM). The goal of Experiment 2 was to test between these hypotheses using fMRI.

Experiment 2

The goal of Experiment 2 was to determine the structural source that accounts for the emerging developmental effect. We employed the paradigm from Experiment 1 as we used fMRI to measure developmental changes among children between the ages of 8 and 15. We also relied on the extant literature to draw conclusions about the role that different brain networks (i.e., inference-making versus ToM) may play in this reading task.

Material and Methods

Participants

Twenty-eight typically developing children participated in this fMRI experiment. All children included in the fMRI study (1) were native French speakers, (2) had no MRI counter-indications, (3) had no history of neurological and psychiatric disorder, (4) had no diagnosis of mental retardation or high intellectual potential and a full-scale IQ between 70 and 130 as

measured by the NEMI-2 (Nouvelle Echelle Metrique de l'Intelligence) (Cognet, 2006), (5) had no diagnosis of autism and a score lower than 32 on the Autism Quotient test (Baron-Cohen et al, 2001), and (6) had no diagnosis of dyslexia and a reading score higher than a standardized score of 70 on the Alouette test [Lefavrais, 1967]). Overall, ten participants were excluded from further analysis for one of the following reasons: performance below or close to 50% on the comprehension questions (n=3), excessive head motion during the imaging session (n=2), poor brain coverage (n=1), technical issues during image acquisition (n=3), or an excessive delay between the imaging session and the completion of the standardized tests (n=1). Thus, the analyses included 18 subjects (9 females), aged from 8.35 to 14.71 (mean=11.61, sd=1.89).

Parents gave their written informed consent and children gave their assent to participate in the experiment. Families were paid 80 euros for their participation. The experiment was approved by the local ethics committee (CPP Lyon Sud-Est II).

Materials

Fully-Deductive and *Implicated-Premise* stories were similar to those used in Experiment 1 (see **Table 1**). However, to maximize signal-to-noise ratio for the fMRI data, 6 additional stories were presented to participants for each story version. That is, there were 24 stories with a disjunctive argument in this experiment (12 *Fully-Deductive* and 12 *Implicated-Premise* stories). Following Prado et al. (2015), 12 filler stories were also included to help disguise the goal of the experiment in the scanner (see **Table 1**). Filler stories did not contain disjunctive arguments.

Experimental procedures

The task and stimulus timing were identical to those in Experiment 1. However, in addition to being read on the screen, each line was spoken over headphones too. Periods of visual fixation between stories constituted the baseline. The fMRI session was split into 6 runs

that contained 6 stories each (2 *Fully-Deductive* stories, 2 *Implicated-Premise* stories, and 2 filler stories). Stories were presented in a pseudo-randomized order, so that two stories of the same type were not following each other within a run. Behavioral responses during scanning were recorded using MR-compatible keypads placed below each hand. Stimuli were generated using Presentation software (Neurobehavioral Systems, Albany, CA) and displayed on a computer screen that was viewed by the participants through a mirror attached to the head coil.

Several days before the fMRI session, participants were invited to the lab to complete the psychometric tests (i.e., IQ, AQ and Alouette reading test) and to practice the task in a mock fMRI scanner. During that practice session, children were presented with 12 filler stories (whose content differed from the filler stories used in the main experiment).

Imaging acquisition

Functional and anatomical images were acquired with a Philips 3T Achieva scanner (Philips Medical systems, Best, Netherlands). A high-resolution anatomical scan was collected for each participant (Field of view=240x240 mm, 512x512 matrix, TR=1278 ms, TE=3.047 ms, flip angle=8°, slice thickness=0.9 mm, number of slices=188). Functional sequences were collected with a gradient-echo, echo-planar sequence (TR=2200 ms, TE=30 ms, flip angle=90°). Twenty-six axial slices were acquired per volume (slice thickness=3.5 mm, Field of view=220 mm, 128x128 matrix).

fMRI preprocessing

Images were analyzed with SPM8 (Statistical Parametric Map 8, Welcome Department of Cognitive Neurology, London, UK). The first 6 volumes of each run were not acquired to allow for T1 equilibration effects. Functional images were corrected for slice acquisition delays and realigned to the first image of the first run to correct for head movements. Volumes with excessive head motion were further identified using the ArtRepair

software (Mazaika et al., 2009). Specifically, following the criteria used in our previous studies (Demir et al, 2014; 2015; 2016; Mathieu et al, 2017), volumes showing rapid scan-to-scan movement greater than 1.5 mm were substituted by the interpolation of the 2 nearest non-repaired volumes. Runs with more than 10% of repaired volumes were excluded from the analysis (based on this cut-off, one run was discarded for 4 participants and 2 runs were discarded for 2 participants). Realigned images were spatially smoothed with a Gaussian filter prior application of the ArtRepair algorithms. Because voxels were anisotropic, this filter was also anisotropic ($4 \times 4 \times 8$ mm full-width at half maximum). Finally, in line with our previous studies on children within the same age range (Prado et al, 2014; Schwartz et al, 2017), all individual brains were normalized into the standard adult MNI space.

Statistical analyses of fMRI data were performed according to the GLM. Following Prado et al. (2015), brain activity elicited by the conclusion of each story was modeled by a block encompassing the last two sentences (i.e., sentences (f) and (g)). The onset of this block was time-locked to the beginning of sentence (f) and the offset was time-locked to the beginning of the question. Sentences (a) to (e) were modeled as events of no interest with an onset time-locked to the beginning of sentence (a) and an offset time-locked to the beginning of sentence (f). The question was modeled as an event of no interest. For each subject and for each run, this model was applied to the three types of stories (*Fully-Deductive*, *Implicated-Premise* and filler stories). Time series were high-pass filtered (1/128 Hertz) and serial correlations were corrected for first-order (AR1) autocorrelations.

fMRI analyses

Following Prado et al. (2015), fMRI data were analyzed using a Region of Interest (ROI) approach. Specifically, we aimed to measure activity in brain networks involved in inference-making and ToM. Several studies have identified the RLPFC, DLPFC and IPL as critical regions for inference-making in children and adults (Crone et al, 2009; Wendelken et

al, 2011; Mathieu et al, 2015; Vendetti & Bunge, 2014). Therefore, we defined bilateral ROIs in each of these regions based on coordinates from our previous study on adult participants that employed the same task (Prado et al., 2015). To define ROIs involved in ToM, we turned to a meta-analysis of neuroimaging studies on ToM (Van Owervalle, 2009). ToM ROIs were located in the bilateral TPJ, the mPFC and the PC (**Table 2**). In line with Prado et al. (2015), all ROIs were 6-mm radius spheres and were defined using the SPM toolbox Marsbar. The fMRI signal (beta weight) was averaged for each subject and story version (*Fully-Deductive* and *Implicated-Premise*) across all voxels within a given ROI. Beta weights from the bilateral RLPFC, bilateral DLPFC and bilateral IPL were then averaged to reflect activity in the inference-making network (see **Fig. 2**). Similarly, beta weights from the bilateral TPJ, mPFC and PC were averaged to reflect activity in the ToM network.

It is important to acknowledge that the sample size in Experiment 2 (n=18) was smaller than in Experiment 1 (n=132). Because there is no previous fMRI study on the development of pragmatic abilities in children, effect size estimates were not available to inform sample size. This raises the possibility that Experiment 2 may be underpowered and that the chance of discovering true effects is relatively low (Button et al., 2013). Therefore, rather than using frequentist statistics as in Experiment 1, we turned to Bayesian analyses to analyze the experiment and assess its sensitivity (Morey et al., 2016; Lee & Wagenmakers, 2013). Specifically, we applied the Sequential Bayes Factors (SBF) procedure to estimate the strength of evidence for age-related changes of activity (H1) versus no age-related changes of activity (H0) as a function of increased sample size, both in the inference-making and the ToM network (Rouder, 2014; Schönbrodt et al, 2017). In the SBF procedure, a Bayes factor (i.e., BF_{10}) is repeatedly calculated as sample size increases, until there is enough evidence for either H1 or H0. Because BF_{10} will always converge either towards 0 (if H0 is true) or towards infinity (if H1 is true), data collection may be stopped when enough evidence is

gathered. As suggested by Jeffreys (1939), a $BF_{10} > 3$ indicates "substantial evidence" for H1 versus H0 (i.e., "substantial" indicating evidence starting to be worth taking note of; Lee and Wagenmakers, 2013). Conversely, a $BF_{10} < 1/3$ indicates substantial evidence for H0 versus H1. Therefore, $3 < BF_{10} < 1/3$ indicates data sensitivity for distinguishing between H1 and H0 and has been used as a stopping rule for data collection in previous studies (e.g., Martin et al., 2016). All analyses were performed with JASP (https://jasp-stats.org). Beta maps for all participants and conditions are available at NeuroVault, along with ROI masks: https://neurovault.org/collections/3040/. Associated behavioral data are available at Mendeley Data: http://dx.doi.org/10.17632/fd7t5h3grb.2. The Matlab scripts used to analyze fMRI data are available upon request.

Results and discussion

fMRI data

We first focused on conclusions of *Implicated-Premise* stories and used Bayesian Pearson correlations to estimate the strength of evidence for age-related change (H1) versus no age-related change (H0) of activity in those stories. In the inference-making network, there was substantial evidence for H1 versus H0 ($BF_{10} = 5.40$), with activity being positively correlated with age ($r = 0.58$) (see **Fig. 3A, left**). The SBF procedure indicated that substantial level of evidence was obtained with about 12 participants (see **Fig. 3A, right**). In the ToM network, however, there was substantial evidence for H0 versus H1 ($r = -0.10$, $BF_{10} = 0.31$) (see **Fig. 3B, left**). This level of evidence was obtained with about 13 participants (see **Fig. 3B, right**). Therefore, there was evidence for a positive relationship between age and activity associated with conclusions of *Implicated-Premise* stories in the inference-making network, but not in the ToM network.

To evaluate how specific these relationships were to *Implicated-Premise* stories, we then used Bayesian Pearson correlations to estimate the strength of evidence for H1 versus H0 in *Fully-Deductive* stories. For conclusions of *Fully-Deductive* stories, we found anecdotal evidence for H0 versus H1 in the inference-making network (r = 0.29, BF_{10} = 0.55) (see **Fig. 4A**), and substantial evidence for H0 versus H1 in the ToM network (r = 0.03, BF_{10} = 0.29) (see **Fig. 4B, left**). The SBF procedure indicated that substantial evidence for H0 versus H1 was obtained with about 15 participants in the ToM network (see **Fig. 4B, right**). Therefore, in the inference-making and ToM networks, there was anecdotal to substantial evidence for a lack of relationship between age and activity associated with conclusions of *Fully-Deductive* stories.

Change of brain activity with age was then directly compared between story versions and networks using a Bayesian ANOVA with default prior scales. The ANOVA included the within-subject factors Story version (*Fully-Deductive* versus *Implicated-Premise*) and Network (inference-making versus ToM), as well as the between-subject factor Age. In a Bayesian ANOVA, different models that include versus do not include each factor and interaction are created[4]. A BF_{10} is then calculated for each model (i.e., reflecting the evidence in favor of that model versus the null model) and Bayesian Model Averaging (BMA) is used to estimate evidence in favor of main effects and interactions (Hoeting et al., 1999). Critically, this analysis revealed substantial evidence for the Story version × Network × Age interaction (BF_{10} = 5). That is, there was substantial evidence for a stronger relationship between activity for *Implicated-Premise* conclusions and age in the inference-making than in the ToM network (correlation between age and difference in activity between inference-making and ToM network for *Implicated-Premise* conclusions: r = 0.59, BF_{10} = 6.20). In contrast, there was anecdotal evidence for an equal relationship between activity for *Fully-Deductive* conclusions

---

[4] Interactions were only included when the component main effects were present (Rouder et al, 2017).

and age in the inference-making and in the ToM network (correlation between age and difference in activity between inference-making and ToM network for *Fully-Deductive* conclusions: r = 0.27, BF_{10} = 0.50) (see **Fig 5**). Evidence for main-effects and other lower-order interactions is shown on **Table 3**. For the sake of completeness, frequentist analyses are described in the Supporting information.

Behavioral data

Even though this was not the main goal of Experiment 2, we also analyzed behavioral responses in the scanner to evaluate whether patterns were similar to those obtained in Experiment 1. First, log-transformed reading times of the conclusion (i.e., sentence (f)) were analyzed in a Bayesian ANOVA with default prior scales. The ANOVA included the within-subject factor Story version (*Fully-Deductive* versus *Implicated-Premise*) and the between-subject factor Age. BMA indicated anecdotal evidence for main effects of Story version (BF_{10} = 2.31) and Age (BF_{10} = 1.31). There was also anecdotal evidence for the Story version×Age interaction (BF_{10} = 1.11). That is, even though a visual examination of conclusion reading times as a function of age and story version suggests an age-related increase in the cost of reading conclusions of *Implicated-Premise* stories when compared to *Fully-Deductive* ones (see **Fig. 6A**), there was only marginal evidence for this pattern with the sample size of Experiment 2. Consistent with an increasing body of literature, this might suggest that neural measures can be more sensitive than behavioral measures in tracking a developmental phenomenon (Gabrieli et al., 2015). Frequentist statistics for reading times of conclusions in Experiment 2 are described in the Supporting information.

Second, responses to justification questions were analyzed using a similar Bayesian ANOVA with default prior scales. BMA indicated extreme evidence for a main effect of condition (BF_{10}> 100). That is, participants were much less likely to accept conclusions in

Implicit-Premise stories than in *Fully-Deductive* stories (24% versus 81%). There was anecdotal evidence for a main effect of Age (BF_{10} = 0.65) and for a Story version × Age interaction (BF_{10} = 2.06). Unlike in Experiment 1, a visual examination of responses to justification questions as a function of age and story version indicates no age-related decrease in *Implicated-Premise* stories (see **Fig. 6B**). That is, unlike in Experiment 1, most young participants responded in an adult-like way to the justification questions of the two main kinds of stories. We can see three explanations for this result. One is that young children, who choose to devote several hours of their time to standardized tests and fMRI scanning, may be more motivated than children carrying out a reading task in school. As a result, they may pay more attention to the difference between the two types of stories. Another possible explanation is that the presence of filler stories in the fMRI experiment (as well as the greater number of trials) allows children to better appreciate the difference between the two main conditions. A last explanation is that Experiment 2 presented the lines of text both visually and aurally. In any case, an examination of developmental changes in reading times for *Implicated-Premise* conclusions (as compared to *Fully-Deductive* ones) points to a pattern that is largely similar to that observed in Experiment 1. So, while future work could address children's superior performance with the off-line justification questions here, indications are that our paradigm is indeed capturing pragmatic inference-making development in the scanner.

General Discussion

In Experiment 1, we found that the reading time of conclusions in the *Implicated-Premise* stories (as compared to *Fully-Deductive* stories) significantly increased with age. Given that pragmatic abilities are well known to develop from childhood to adolescence (e.g. Huang & Snedeker, 2009; Katsos et al., 2016; Pouscoulous et al., 2007), this effect may be

due to the fact that a conclusion in the *Implicated-Premise* condition can be considered as a source for an enriching pragmatic inference. In Experiment 2, we used fMRI to investigate the neural bases of this developmental trend in children from 8 to 15. We began with two competing hypotheses. One was that the emergence of pragmatic inference-making would be linked to areas known for other sorts of inference-making; the other was that the emergence of pragmatic inference-making would be linked to the development of ToM structures. We found evidence for age-related increases of activity in fronto-parietal regions associated with inference-making, but not in regions associated with ToM.

Pragmatic processing of conclusions in Implicated-Premise stories increases with age

At the behavioral level (Experiment 1), our results indicate that the reading time of conclusions in *Implicated-Premise* stories -- as compared to *Fully-Deductive* stories -- becomes increasingly distinctive between middle childhood and adulthood. This can be seen in two ways. First, our oldest participants took significantly longer to read the conclusions in *Implicated-Premise* stories as opposed to the conclusions in *Fully-Deductive* ones. Second, the acceptance rates of justification questions (asking whether the protagonist was justified to draw the conclusion) dropped progressively with age and significantly so for the *Implicated-Premise* stories when compared to the *Fully-Deductive* ones. Interestingly, Experiment 1's younger participants' rates of assenting to the justification questions are comparable across the *Fully-Deductive* and *Implicated-Premise* stories. The results among the older participants are consistent with our previous adult study (Prado et al, 2015). Overall, indications are that, with age, reading conclusions in the *Implicated-Premise* condition requires extra-processing as compared to conclusions in the *Fully-Deductive* condition.

Increased pragmatic processing of conclusions in Implicated-Premise stories is associated with increased activity in the inference-making network

We now turn to our main question (*what develops with the more demanding pragmatic task?*) by considering participants' neural activity associated with processing conclusions in *Implicated-Premise* and *Fully-Deductive* stories (Experiment 2). Using Bayesian analyses, we found substantial evidence for a lack of age-related increase of activity in the ToM network, both for conclusions of *Implicated-Premise* and *Fully-Deductive* stories. That is, changes of activity in the ToM network are unlikely to account for the developmental trend observed in Experiment 1. In the inference-making network, however, we found evidence for age-related increases of activity that were specific to *Implicated-Premise* conclusions and to that network. That is, these changes were (i) not observed for *Fully-Deductive* conclusions and (ii) larger in the inference-making than in the ToM network. The inference-making network, which includes the RLPFC, DLPFC and IPL, has been consistently identified in various studies of conditional reasoning (Monti et al., 2007; Noveck et al., 2004), disjunctive reasoning (Reverberi et al, 2007), categorical reasoning (Goel et al., 2001; Prado et al., 2013), transitive reasoning (Fangmeier et al., 2006; Prado et al., 2008; 2013), matrix reasoning (Crone et al., 2009) and analogical reasoning (Wendelken et al., 2008). It has been argued that the different regions composing this system may be involved in different aspects of inference-making. For example, whereas the IPL and the DLPFC might be involved in the representation and maintenance of information (e.g., premises) in working memory (Badre & D'Esposito, 2007), the RLPFC might support relational integration and play a critical role in integrating premises of arguments in order to produce or evaluate conclusions (Bunge et al., 2009; Prado et al, 2013; Vendetti and Bunge, 2014). Overall, the hypothesized developmental emergence of pragmatic inference-making in our task is more likely associated with the maturation of the inference-making network than with greater engagement of the ToM network.

Relevance of our findings to the literature on the neural bases of pragmatic processing

We believe that the current findings speak to recent neuroimaging findings on scalar inferences (which, as pointed out earlier, concerns the way an underinformative utterance such as *Some x are y* can be enriched to mean *Some but not all x are y*), arguably the most investigated pragmatic inference in the cognitive sciences. A recent line of neuroimaging work (Shetreet et al. 2014a, 2014b; 2014c), which includes a developmental study, presents data supporting the claim that the main site responsible for scalar inference-making ought to be found in language areas (viz. left Inferior Frontal Gyrus); this claim stems from a theoretical approach that treats scalar inferences as linguistic in nature (Chierchia et al., 2004). Yet, this region is not part of the inference-making network in which we found age-related increases of activity in our task. Given the current state of the literature on inference-making (which covers logical, linguistic, and pragmatic inference), we briefly consider two reasons to question the role of linguistic mechanisms from the left Inferior Frontal Gyrus in pragmatic inference-making. First, the part of the left IFG that was found activated in Shetreet et al. in 2014 (i.e., Brodmann Area [BA] 47) does not subserve linguistic functioning exclusively. For example, Monti et al. (2009) showed – through the juxtaposition of two very similar tasks that aimed to determine the neural underpinnings of logical inference-making and of linguistic inference-making – that BA47 is implicated in both linguistic inferencing *and* logical inferencing. This suggests that activity in BA47 does not necessarily point to linguistic computations only. Second, it is interesting to note that Shetreet et al.'s most discerning isolating case (one which compares an underinformative *Some* statement to their task's two other *Some* statements) points to BA10, i.e., a region of the RLPFC that is part of the inference-making network in the present task. Thus, a stronger claim for the literature, which encompasses Shetreet et al.'s data, is arguably that regions of the inference-making

network as defined here – including the RLPFC – plays a key role in pragmatic inference-making.

Limitations

We consider two potential limitations to our work. First, we did not find any evidence for a developmental increase of activity associated with *Implicated-Premise* stories in the ToM network. Based on that observation and on evidence that activity increased in the inference-making network, we concluded that the development of pragmatic inference-making in our task was more likely due to the maturation of the inference-making than the ToM network. However, it is important to note that all children in the present study were older than 8. Thus, it remains possible that developmental changes of activity in ToM regions may underlie the development of pragmatic processing in younger children. In other words, although the present study suggests that the development of ToM mechanisms is unlikely to account for the development of pragmatic inference-making from 8 to 14, future research might investigate to what extent it might for younger children.

Second, performing cross-sectional fMRI experiments is challenging and the final sample size of Experiment 2 (n=18) is limited, especially when compared to that of Experiment 1 (n=132). It is therefore possible that Experiment 2 is not sensitive enough to detect true effects. To assess the sensitivity of our fMRI analyses, Experiment 2 was analyzed using Bayesian rather than frequentist statistics (Morey et al., 2016; Lee & Wagenmakers, 2013). That is, we reported the strength of evidence for H1 (there is an age-related change of activity) versus H0 (there is no age-related change of activity) in all analyses, thereby allowing readers to quantify evidence for the effects observed. We also used the Sequential Bayes Factors (SBF) procedure to indicate the number of participants required to reach substantial evidence for H1 versus H0 (and vice versa) (Rouder, 2014; Schönbrodt et al,

2017). Although we found substantial evidence for most of the critical effects with our sample size), we acknowledge that this sample size remains small. Thus, these fMRI results should be seen as providing the groundwork for future studies that might investigate the development of pragmatic inference-making in fMRI studies with larger sample sizes.

Conclusion

In sum, we investigated the neural development of pragmatic inference-making in natural discourse. While using ecologically valid texts that reflect natural discourse, our behavioral study is in keeping with the idea that pragmatic inferencing may increase with age through adolescence. At the neural level, we did not find evidence for age-related changes of activity in the ToM network. Instead, we found evidence indicating that activity in the inference-making network significantly increased from childhood through adolescence. Thus, the hypothesized development of pragmatic inference-making in our task is more likely supported by a functional maturation of inference-making brain structures rather than a maturation of ToM areas.

References

Ashburner, J., Friston, K. J. (2005). Unified segmentation. Neuroimage, 26, 839-851.

Badre, D. and D'Esposito, M. (2007). Functional Magnetic Reasonance Imaging evidence for a hierarchical gradient in the Prefrontal cortex. Journal of Cognitive Neuroscience, 19(12), 2082-2099.

Baillargeon, R., Scott, R.M. and He, Z. False-belief understanding in infants (2010). Trends in Cognitive Sciences, 14(3), 110–118.

Barnea-Gorali, N., Menon, V., Eckert, M., Tamm, L., Bammer, R., Karchemskiy, A., Dant, C.T., Reiss, A.L. (2005). White matter development during childhood and adolescence: a cross-sectional diffusion tensor imaging study, Cerebral Cortex, 15, 1848-1854.

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

Brainard, D.H. (1997). The psychophysics toolbox. Spatial Vision, 10, 433–436.

Bunge, S.A., Helskog, E.H. and Wendelken, C. (2009) Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. NeuroImage, 46, 338-342.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafo, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci, 14*(5), 365-376. doi: 10.1038/nrn3475

Caillies, S. and Le Sourn-Bissaoui, S. (2008). Children's understanding of idioms and theory of mind development. Developmental Science, 11, 5, 703-711.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Structures and Beyond (Belletti, A., ed.), pp. 39–103, Oxford University Press

Cognet, G. (2006). Nouvelle Echelle Métrique de l'Intelligence. Editions du Centre de Psychologie Appliquée, Paris.

Coulson, S. & Van Petten, C. (2002). Conceptual integration and metaphor: an event-related potential study. Memory & Cognition, 30(6), 958-968.

Crone, E.A., Wendelken, C., van Leijenhorst, L., Honomichl, R.D., Christoff, K. and Bunge, S.A. (2009) Neurocognitive development of relational reasoning. Developmental Science, 12, 55-66.

Demir, O. E., Prado, J., & Booth, J. R. (2014). The differential role of verbal and spatial working memory in the neural basis of arithmetic. *Dev Neuropsychol, 39*(6), 440-458. doi: 10.1080/87565641.2014.939182

Demir, O. E., Prado, J., & Booth, J. R. (2015). Parental socioeconomic status and the neural basis of arithmetic: differential relations to verbal and visuo-spatial representations. *Dev Sci, 18*(5)*,* 799-814. doi: 10.1111/desc.12268

Demir, O.E., Prado, J., & Booth, J.R. (2016). Neural correlates of math gains vary depending on parental socioeconomic status (SES). *Frontiers in Psychology, 7*, 892.

Dumontheil, I., Apperly, I. A. and Blakemore, S. J. (2010a) Online usage of theory of mind continues to develop in late adolescence. Developmental Science, 13(2), 331-338.

Dumontheil, I., Houlton, R., Christoff, K. and Blakemore, S.J. (2010) Development of relational reasoning during adolescence. Developmental Science, 13(1), 15-24.

Dumontheil, I. (2014) Development of abstract thinking during childhood and adolescence: The role of rostrolateral prefrontal cortex. Developmental Cognitive Neuroscience, 10, 57-76.

Epley, N., Morewedge, C.K., & Keysar, B. (2004). Perspective taking in children and adults: equivalent egocentrism but differential correction. Journal of Experimental Social Psychology, 40, 760–768.

Fangmeier, T., Knauff, M., Ruff, C. C., & Sloutsky, V. (2006). FMRI evidence for a three-stage model of deductive reasoning. Journal of Cognitive Neuroscience, 18, 320–334.

Filippova, E. and Astington, W. (2010) Children's understanding of Social-Cognitive and Social-Communicative Aspects of Discourse Irony. Child Development, 81(3), 913-928.

Gabrieli, J. D., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron, 85*(1), 11-26. doi: 10.1016/j.neuron.2014.10.047

Goel, V. and Dolan, R.J. (2001). Functional neuroanatomy of three-term relational reasoning. Neuropsychologia, 39, 901-909.

Grice, P. (1989) Studies in the Way of Words, Harvard University Press

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. Statistical Science, 14, 382–417.

Huang, Y., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in five-year olds: Evidence from real time spoken language comprehension. Developmental Psychology, 45, 1723–1739.

Jeffreys H. (1939). *The Theory of Probability*, 1st/3rd Edn. Oxford: Oxford University Press.

Josephs, O., Turner, R. and Friston, K. (1997) Event-related fMRI. Human Brain Mapping, 5, 243-248.

Katsos, N., Cummins, C., Ezeizabarrena, M., Gavarro, A., Kraljevic, J.K. et al. (2016). Cross-linguistic patterns in the acquisition of quantifiers. Proceedings of the National Academy of Science of the United States of America, 113(33), 9244-9249.

Keysar, B., Barr, D.J., Balin, J.A., & Brauner, J.S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. Psychological Science, 11, 32–38.

Keysar, B., Lin, S., & Barr, D.J. (2003). Limits on theory of mind use in adults. Cognition, 89, 25–41.

Kissine, M. (2015). Pragmatics as metacognitive control. Frontiers in Psychology, 6, 2057.

Kutas, M. & Hillyard,S.A. (1984) Brain potentials during reading reflect word expectancy and semantic associtation. Nature, 307(5947), 161-163.

Lee M. D., Wagenmakers E. J. (2013). Bayesian Cognitive Modeling: A Practical Course. Cambridge, MA: Cambridge University Press

Lefavrais, P. (1967) Test de l'Alouette, Éditions du Centre de Psychologie Appliquée, Paris (2ème ed.)

Mathieu, R., Booth, J. and Prado, J. (2015) Distributed neural representation of logical arguments in school-aged children. Human Brain Mapping, 36, 996-1009.

Mathieu, R., Epinat-Duclos, J., Leone, J., Fayol, M., Thevenot, C., & Prado, J. (2017). Hippocampal spatial mechanisms relate to the development of arithmetic symbol processing in children. *Dev Cogn Neurosci*. doi: 10.1016/j.dcn.2017.06.001

Martin, J. R., Sackur, J., Anlló, H., Naish, P., & Dienes, Z. (2016). Perceiving time differences when you should not: Applying the El Greco fallacy to hypnotic time distortions. *Frontiers in psychology*, 7.

Mazaika, P.K., Hoeft, F., Glover, G.H. and Reiss, A.L. (2009) Methods and Software for fMRI Analysis of Clinical Subjects. Neuroimage, 47: S58.

Mills, K.L., Lalonde, F., Clasen, L.S., Giedd, J.N., Blakemore, S.J. (2014) Developmental changes in the structure of the social brain in late childhood and adolescence. Social Cognitive and Affective Neuroscience, 9(1), 123-131.

Morey, R. D., Romeijn, J. W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, 72, 6-18.

Monti, M. M., Osherson, D. N., Martinez, M. J., & Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: A language-independent distributed network. Neuroimage, 37, 1005–1016.

Monti, M.M., Parsons, L.M. and Osherson, D.N. (2009) The boundaries of language and thought in deductive inference. P. Natl. Acad. Sci, 106, 12554-12559.

Norbury C.F. (2014) Practitioner review : social (pragmatic) communication disorder conceptualization, evidence and clinical implications. Journal of Child Psychology and Psychiatry, 55(3), 204-216.

Noveck, I. A. (2001) When children are more logical than adults: Experimental investigations of scalar implicature. Cognition, 78(2), 165-188.

Noveck, I. A., Goel, V., & Smith, K. W. (2004). The neural basis of conditional reasoning with arbitrary content. Cortex, 40, 613–622.

Noveck, I. A. and Reboul, A. (2008). Experimental pragmatics: a Gricean turn in the study of language. Trends in Cognitive Sciences, 12(11), 425-431.

Noveck, I.A., Bonnefond, M. and Van der Henst, J.B. (2011). Squib: A deflationary account of invited inferences. Belgian journal of linguistics, 25(1), 195-208

Olfield, R.C. (1971) The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9, 97-113.

Papafragou, A. & Skordos, D. (in press). Scalar Implicature. In J. Lidz, W. Snyder and J. Pater (eds.), The Oxford Handbook of Developmental Linguistics. Oxford: Oxford University Press.

Politzer-Ahles, S. and Gwilliams, L. (2015) Involvement of prefrontal cortex in scalar implicatures : evidence from magnetoencephalography . Language, Cognition and Neuroscience, 30(7), 853-866.

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. Language Acquisition, 14, 347–375.

Prado, J., Chadha, A. and Booth, J.R. (2011) The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. Journal of Cognitive Neuroscience, 23, 3483-3497.

Prado, J., Mutreja, R., & Booth, J. R. (2013). Fractionating the neural substrates of transitive reasoning: Task-dependent contributions of spatial and verbal representations. Cerebral Cortex, 23, 499–507.

Prado, J., Mutreja, R., & Booth, J. R. (2014). Developmental dissociation in the neural responses to simple multiplication and subtraction problems. *Dev Sci, 17*(4), 537-552.

Prado, J., Spotorno, N., Koun, E., Hewitt, E., Van der Henst, J.B., Sperber, D.and Noveck, I.A. (2015) Neural interaction between logical reasoning and pragmatic processing in narrative discourse, Journal of Cognitive Neuroscience, 27(4), 692-704.

Prado J., Van der Henst J.B. and Noveck I.A. (2008) Spatial associations in relational reasoning: evidencefor a SNARC-like effect. Q J Exp Psychol, 61,1143-1150.

Reverberi, C., Cherubini, P., Frackowiak, R. S. J., Caltagirone, C., Paulesu, E., & Macaluso, E. (2010). Conditional and syllogistic deductive tasks dissociate functionally during premise integration. Human Brain Mapping, 31, 1430–1445.

Reverberi, C., Cherubini, P., Rapisarda, A., Rigamonti, E., Caltagirone, C., Frackowiak,R.S.J., Macaluso, E. and Pauluso, E. (2007) Neural basis of generation of conclusion in elementary deduction. Neuroimage, 38, 752-762.

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301-308.

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E. J. (2017). Bayesian analysis of factorial designs. Psychological Methods, 22(2), 304.

Rumain, B., Connell, J. and Braine, M. D. (1983) Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. Developmental Psychology, 19(4), 471-481.

Saxe, R.R., Whitfield-Gabrieli, S., Scholz, J., Pelphrey, K.A. (2009) Brain regions for perceiving and reasoning about others in school-aged children, Child Development, 80(4), 1197-1209.

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322.

Schwartz, F., Epinat-Duclos, J., Leone, J., & Prado, J. (2017). The neural development of conditional reasoning in children: Different mechanisms for assessing the logical validity and likelihood of conclusions. *Neuroimage, 163*, 264-275. doi: 10.1016/j.neuroimage.2017.09.029

Shanahan, T., Kamil, M. L., & Tobin, A. (1982). Cloze and intersentential comprehension. Reading Research Quarterly, 17, 229-255.

Shetreet, E., Chierchia, G., & Gaab, N. (2014a). When some is not every: Dissociating scalar implicature generation and mismatch. Human Brain Mapping, 35, 1503–1514.

Shetreet, E., Chierchia, G., & Gaab, N. (2014b). When three is not some: On the pragmatics of numerals. Journal of Cognitive Neuroscience, 26, 854–863.

Shetreet, E., Chierchia, G., & Gaab, N. (2014c). Linguistic inability or poor performance: dissociating scalar implicature generation and mismatch in the developing brain, Developmental Psychology, 50(9), 2264-2275

Southgate, V., Senju, A. and Csibra, G. (2007) Action anticipation through attribution of false-belief by 2-year-olds. Psychological Science, 18(7), 587-592.

Sperber, D. and Wilson, D. (2002) Pragmatics, Modularity and Mindreading. Mind Lang. 17, 3–23.

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25-39.

Van Overwalle, F. (2009) Social cognition and the brain: a meta-analysis. Human Brain Mapping. 30 (3), 829–858.

Vendetti, M.S., Bunge, S.A. (2014) Evolutionary and Developmental Changes in the Lateral Frontoparietal Network: A Little Goes a Long Way for Higher-Level Cognition, Neuron, 84, 906-917.

Wellman, H.M., Cross, D. and Watson, J. (2001) Meta-analysis of theory-of-mind development: the truth about false belief. Child Development, 72(3), 655-684.

Wendelken, C., Nakhabenko, D., Donohue, S.E., Carter, C.S., and Bunge, S.A. (2008). ''Brain is to thought as stomach is to?? '': investigating the role of rostrolateral prefrontal cortex in relational reasoning. Journal of Cognitive Neuroscience, 20, 682–693.

Wendelken, C., O'Hare, E.D., Whitaker, K.J., Ferrer, E. and Bunge, S.A. (2011) Increased selectivity over development in rostrolateral prefrontal cortex. Journal of Neuroscience, 31(47), 17260-17268.

Westlye, L.T., Walhovd, K. B., Dale, A.M., Bjornerud, A., Due-Tonnessen, P., Engvig, A. et al. (2010) Differentiating maturational and aging-related changes of the cerebral cortex by use of thickness and signal intensity. Neuroimage, 52, 172-185.

Figure legends

Figure 1

Results from Experiment 1. (A) Mean reading time of the conclusion as a function of age group and story version. (B) Mean acceptance rate for the justification question as a function of age group and story version.

Figure 2

Location of ROIs constituting the inference-making and the ToM networks on coronal slices of an MRI-normalized brain.

Figure 3

Age-related changes in activity associated with conclusions of *Implicated-Premise* stories. (A) Correlation between activity and age (left) and associated Bayes factor as a function of increased sample size (i.e., Sequential Bayes Factors analysis) (right). (B) Correlation between activity and age (left) and associated Bayes factor as a function of increased sample size (right) in the ToM network.

Figure 4

Age-related changes in activity associated with conclusions of *Fully-Deductive* stories. (A) Correlation between activity and age (left) and associated Bayes factor as a function of increased sample size (i.e., Sequential Bayes Factors analysis) (right). (B) Correlation between activity and age (left) and associated Bayes factor as a function of increased sample size (right) in the ToM network.

Figure 5

Correlations between age and difference in activity between inference-making and ToM networks for *Implicated-Premise* and *Fully-Deductive* conclusions.

Figure 6

Behavioral results from Experiment 2. (A) Mean reading time of the conclusion as a function of age and story version. (B) Mean acceptance rate for the justification question as a function of age and story version.
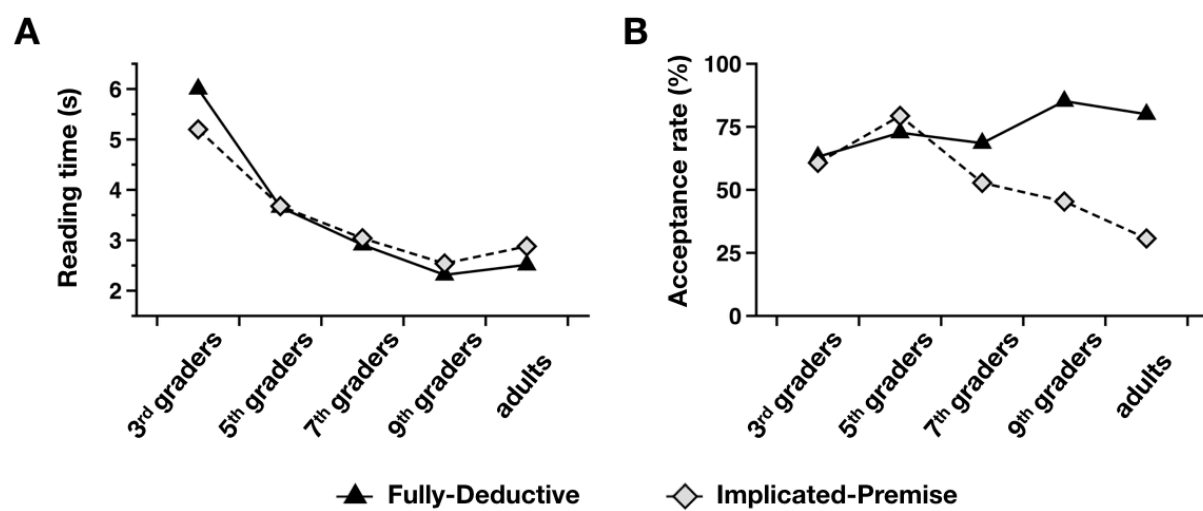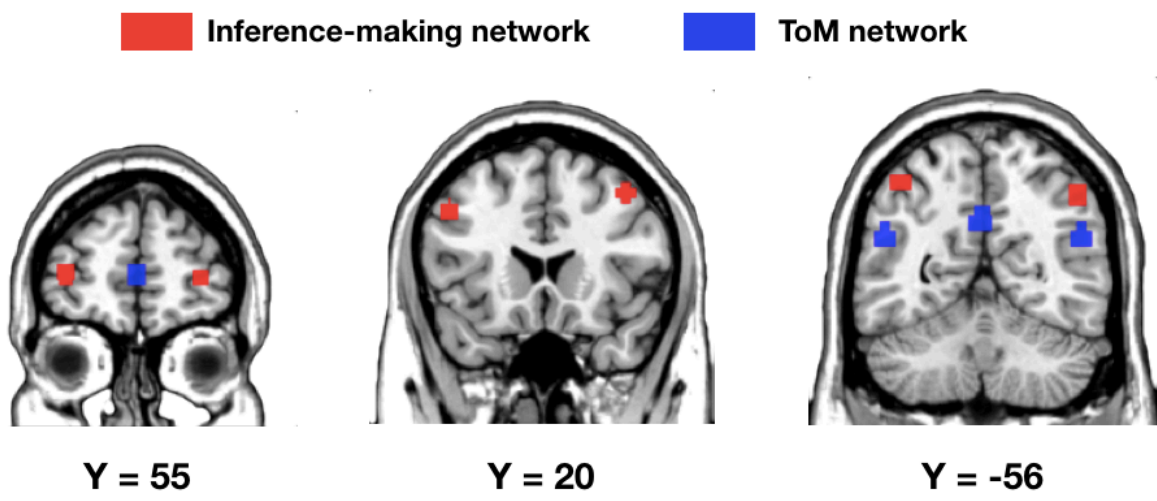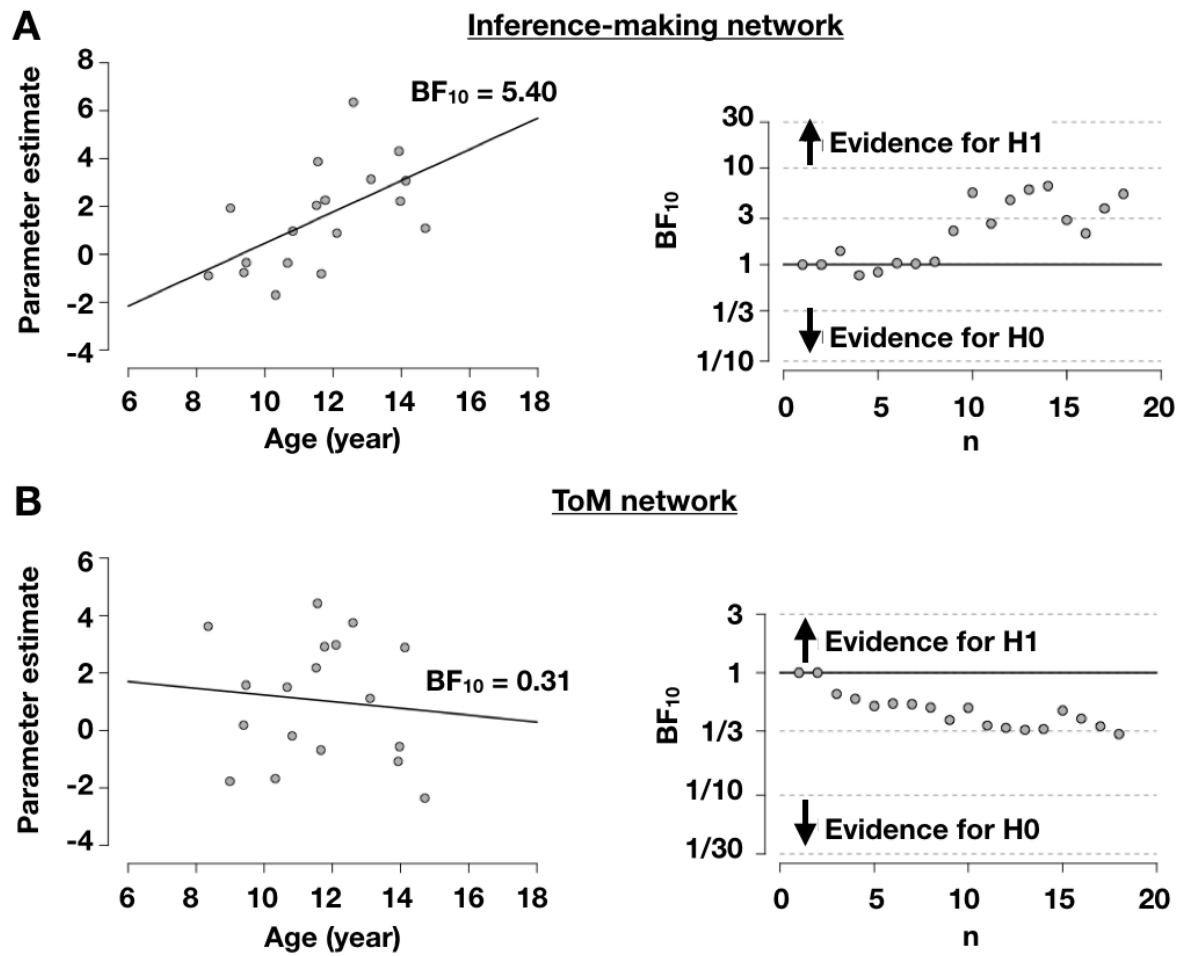
A

Reading time (s)

3rd graders 5th graders 7th graders 9th graders adults

B

Acceptance rate (%)

3rd graders 5th graders 7th graders 9th graders adults

▲ Fully-Deductive   ◇ Implicated-Premise

Figure 1

Inference-making network **ToM network**

Y = 55   Y = 20   Y = -56

Figure 2

A Inference-making network

B ToM network

Figure 3

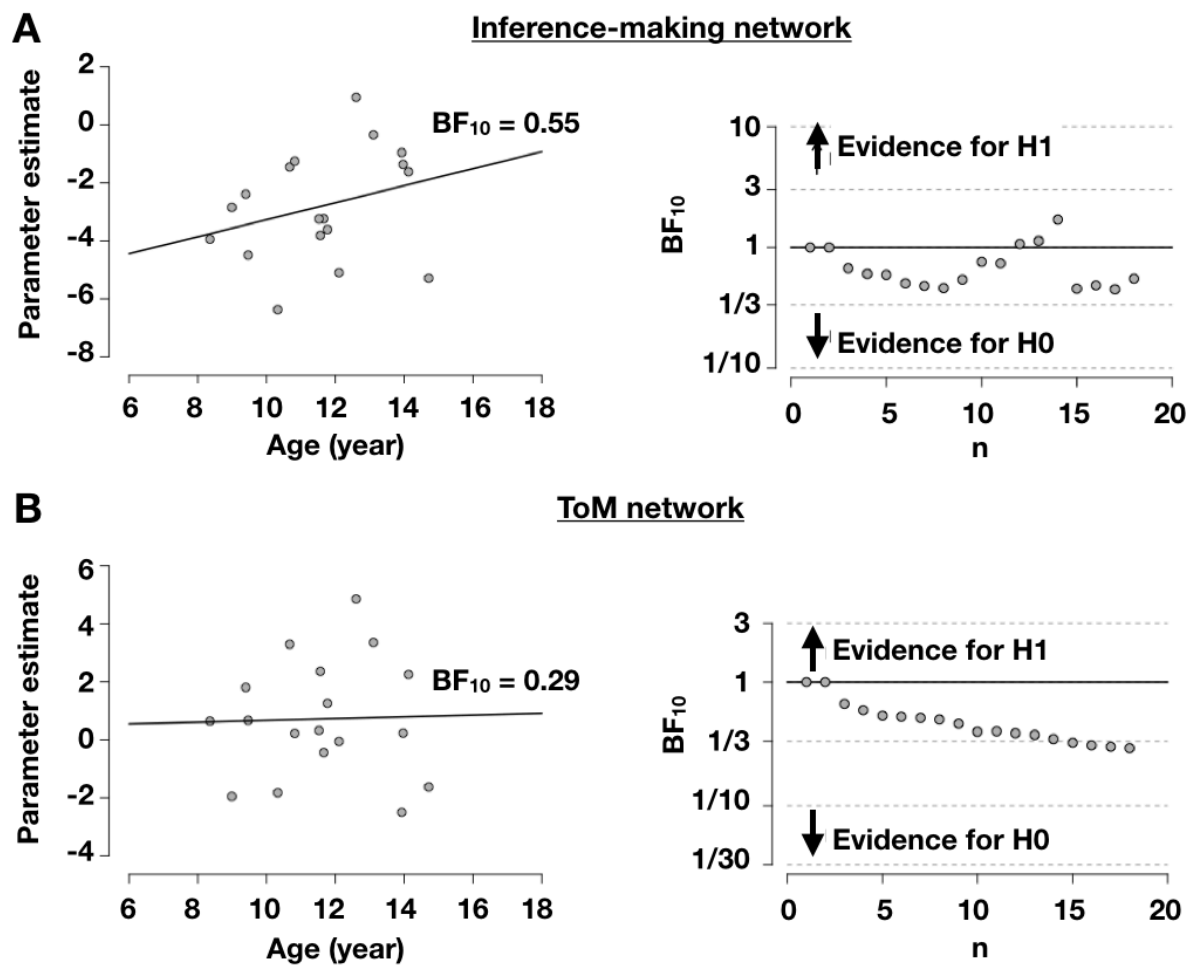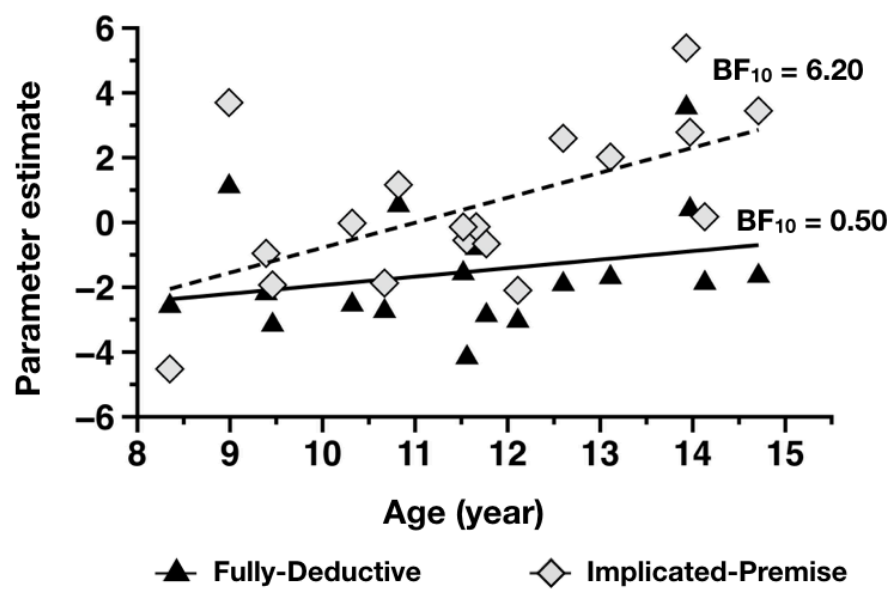**A** Inference-making network

B ToM network

Figure 4

Figure 5

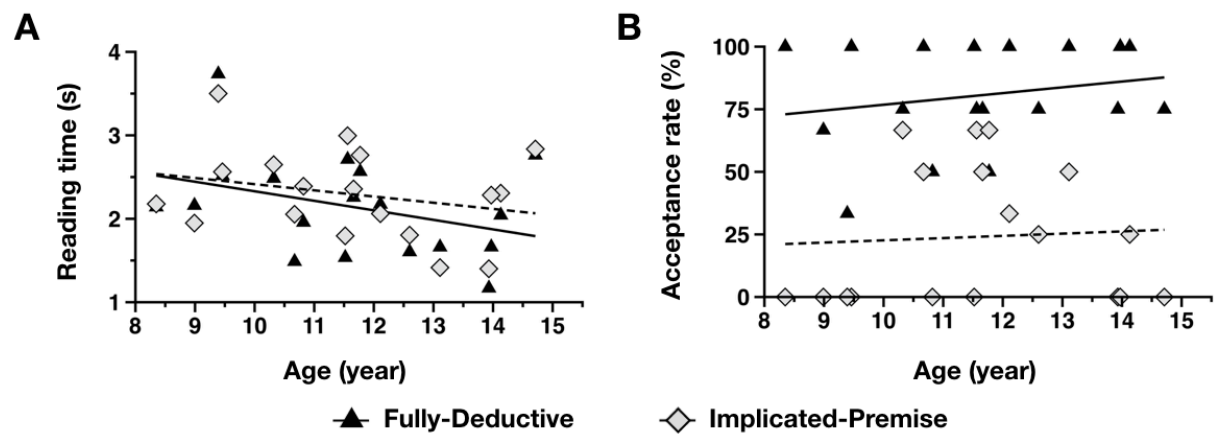**Figure 6**

Table 1. Two examples of disjunctive stories and a filler story used in the experiments (translated from French).

Disjunctive story 1	a. Xavier wants to surprise Claire by inviting her to a restaurant.
Fully-Deductive cond: Implicated-Premise cond:	b. He hesitates between these three days of the week: c. Thursday, Friday or Saturday. d. Claire cannot go on Saturday because she has a theatre course. e. Moreover, Thursday would not work because he is having dinner with his parents. e'. Moreover, Saturday would not work because he is having dinner with his parents. f. He says: "I'll invite her for Friday." g. He hopes she will be available that day.
Comprehension question:	*In your opinion, does Xavier plan to invite his girlfriend?*
Justification question:	*In your opinion, is Xavier justified to think that he should invite Claire on Friday?*

Disjunctive story 2	a. Michael wants to start playing a musical instrument.
Fully-Deductive cond: Implicated-Premise cond:	b. He hesitates between these three instruments: c. The drums, the piano and the guitar. d. His dad does not want him to play (take classes in) drums because it is too noisy. e. Further, he cannot play (take classes in) piano because it is way too expensive. e'. Further, he cannot play (take classes in) drums because it is way too expensive. f. He says: "I will start taking (classes on) the guitar." g. He hopes he will improve quickly.
Comprehension question:	*In your opinion, does Michael want to play a music instrument?*
Justification question:	*In your opinion, is Michael justified to think that he has to start playing the guitar?*

Filler story	a. Jennifer just bought a new camera.
	b. She wants to unpack it immediately. c. It would be easier with a sharp knife. d. She goes to the kitchen to pick up a sharp knife. e. After a few minutes, she retrieves the camera from the package. f. She decides to read the instructions before using it. g. She is very excited to start taking pictures.
Comprehension question:	*In your opinion, is Jennifer happy with what she bought?*

Notes. Filler stories were presented only in Experiment 2 and were always followed by a comprehension question. Otherwise, two thirds of the Fully-Deductive and Implicated-Premise stories were followed by a comprehension question and one third by a justification question (see text).

Table 2. MNI Coordinates of ROIs from the inference-making and ToM networks

Anatomical Location	~BA	MNI Coordinates		
Inference-making network				
L. RLPFC	10	-36	56	7
L. DLPFC	9	-48	23	37
L. IPL	40	-42	-58	52
R. RLPFC	10	33	59	4
R. DLPFC	9	42	17	46
R. IPL	40	48	-55	46
ToM network				
L. TPJ	40	-50	-55	25
R. TPJ	40	50	-55	25
mPFC	9	0	55	6
PC	7	-1	-56	33

L. = left; R. = right; ~BA = approximate Brodmann's area.

Table 3. Analysis of effects for the Bayesian ANOVA with the effects Story version (Implicated-Premise versus Fully-Deductive), Network (Inference-making versus ToM) and Age.

Effects	P(incl)	P(incl|data)	BF Inclusion
Story version	0.737	1.000	2.957e +10
Network	0.737	0.983	20.552
Age	0.737	0.955	7.553
Story version × Network	0.316	0.948	39.833
Story version × Age	0.316	0.639	3.829
Network × Age	0.316	0.914	22.923
Story version × Network × Age	0.053	0.218	5.005