UNIVERSITAT de
BARCELONA

Developmental Intergroup Theory of Mind:
The protagonist's social group membership
as a pragmatic performance factor

Carlota Saumell Andreu

Developmental Intergroup Theory of Mind:

The protagonist's social group membership as a pragmatic performance factor.

Carlota Saumell Andreu

TESI DOCTORAL Universitat de Barcelona 2022

Thesis supervisors:
Dr. Ferran Pons
Dr. Mireia Hernández

Department of Cognition, Development and Educational Psychology

Cognitive Science and Language PhD Programme

"From now on," he said, "I don't want anyone to come in to see me while I'm here. Is that clear?

"Yes, sir," said Sergeant Towser. "Does that include me?"

"Yes."

"I see. Will that be all?"

"Yes."

"What shall I say to the people who do come to see you while you're here?"

"Tell them I'm in and ask them to wait."

"Yes, sir. For how long?"

"Until I've left."

"And then what shall I do with them?"

"I don't care."

"May I send them in to see you after you've left?"

"Yes."

(Heller, 1994, p.129)

To the alpacas.

Acknowledgments

Estic molt agraïda a tothom que ha fet possible aquesta tesi. Al meus supervisors, el Ferran i la Mireia, per la seva confiança i la seva flexibilitat. Per haver-me deixat volar i haver reedreçat els meus pardals quan tocava. Al Joan per marcar el camí i ajudar-me en tot i més.

A tots els patufets i patufetes que seien a la cadira i m'escoltaven pacientment. O ignoraven completament el meu estudi i m'explicaven teories molt més interessants. O m'ensenyaven els peus dels seus germans per Zoom. Gràcies també a totes les famílies que han permès que els seus fills participessin.

I am incredibly grateful to Yarrow Dunham for his extreme generosity and impeccable guidance. It's been an honor to be part of such a thought-provoking research group. I want to thank the lab managers, the fellow PhD students, and the research assistants for welcoming me into the team with open arms.

A les teresines i els meus amics d'aquí i d'allà per amenitzar aquests anys amb molt menjar i molta xerinola. To Lucas for sharing the pain and the glory and for never having to proofread this again.

A la familieta per tot el suport i l'amor. A la meva mare per l'energia i la màgia, al meu pare per la responsabilitat i la curiositat, a la Muckie(+1) per la complicitat i l'humor i a l'Álvaro i a la Pia per la innocència. I al Pablo, que encara no es creu que hagi fet la tesi.

Al Ruchie del Van per estimar-me i fer-me mare. Per cuidar el Refugi amb pantalons de swagger. Al meu escamarlà, la Blau, per animar-me des de la panxulina i per inspirar-me i fer-me créixer cada dia. I ajudar-me a relativitzar les hores de son.

Per últim, al camió que em va atropellar i em va fer replantejar si volia seguir treballant a una agència de publicitat.

Abstract

During the early preschool years, the development of the concept of "the self" and "the other" has a significant impact on two socio-cognitive processes (Theory of Mind and social categorization) that have largely been studied separately. Developing an understanding that there is a "me" and a "you" is crucial to infer someone's mental states (desires, beliefs, and intentions), which is an indispensable skill for interpersonal relationships. However, when these early developing categorization processes extrapolate the individual distinction ("me" and "you") to a group distinction ("us" versus "them"), the interpersonal relationships become biased with an indiscriminate preference for those that belong to one's group. While categorizing is necessary for infants to represent the world in a less taxing manner (Dunham & Olson, 2016) it often triggers stereotypical representations of individuals, not attending to who they are but rather which group they belong to.

Children's performance in false belief (FB) paradigms has been the subject of a heated debate over the last thirty years. The systematic failure of the standard FB tasks before the 4th birthday has been interpreted in divergent manners: nativist accounts attribute it to the high processing demands of the task in relation to children's underdeveloped executive and attentional resources before that age (e.g., Baillargeon et al., 2010; Leslie, 2005; Leslie et al., 2004; Leslie et al., 2005; Scott et al., 2010); empiricist accounts attribute the failure to a lack of a meta-representational concept of belief (Wellman, 2001; Gopnik and Wellman, 1992) and pragmatic accounts attribute children's poor performance to the specific contextual elements of the task that make it pragmatically difficult for children (Helming et al., 2016; Rubio-Fernández & Geurts, 2013; Siegal and Beattie, 1991; Westra, 2017).

The Outgroup Advantage Pragmatic Account (OAPA) presented in this dissertation is a pragmatic account that attempts to explain children's difficulty in the FB task through the lens of the cognitive branch of Developmental Intergroup Theory of Mind. The focus of the account is to show that performance in standard FB tasks is not solely linked to children's competence in attributing false beliefs but rather it is interwoven with pragmatic factors that make the task cognitively demanding. The OAPA establishes that the social group membership of the recipient of the mentalization (the "false-belief holder") is a pragmatic factor that can influence children's performance in a task that directly pits the protagonist's perspective against that of the participant. It predicts that an outgroup protagonist reduces the pragmatic demands of the FB task because it presents an easier target to attribute false beliefs to for three reasons: an outgroup protagonist increases the saliency of the protagonist's perspective, it increases the protagonist's expected ignorance, and it clarifies the experimenter's intent.

The experimental work was divided into four sections. In the first section I explored the pragmatic influence of manipulating the protagonist's social group membership with a real social category (race) and a MGP manipulation in 4-to-5-year-olds' FB performance. I found

that an outgroup protagonist yielded a better performance in the unexpected-contents FB task (Study 1) and influenced the sequential pattern of difficulty of the well-established five-item ToM scale by Wellman and Liu (2004) (Studies 1 and 2).

The second section was designed to address the account's first pragmatic factor: the saliency of the protagonist's perspective. Studies 3 and 4 provided confirmation that manipulating the saliency of the protagonist's perspective influenced 4-to-5-year-olds' FB performance in the unexpected-contents task.

The third section was built to test the account's second pragmatic factor: the protagonist's expected ignorance. Study 5 demonstrated that a shared-group membership between protagonist and agent made the change-of-location FB task more cognitively demanding for 5-to-6-years-olds.

The fourth and last section aimed to test the developmental trajectory of the outgroup advantage in a continuous change-of-location FB task with a population that already mastered standard FB tasks. Testing 8-to-9-year-olds and adults did not show evidence for an outgroup advantage.

The summation of the results of the seven studies presented in this dissertation provided confirmative evidence of the OAPA in 4-to-6-year-olds' performance in FB tasks. The OAPA contributes to a more flexible understanding between competence and performance in children's mentalizing skills in FB task and suggests a path to develop intervention strategies to prevent children's intergroup biases.

Resum

Durant els primers anys d'educació infantil, el desenvolupament del concepte "el jo" i "l'altre" té un impacte significatiu en dos processos sociocognitius (Teoria de la Ment i categorització social) que majoritàriament s'han estudiat per separat. Comprendre la diferència entre un "jo" i un "tu" és crucial per inferir els estats mentals d'algú (desitjos, creences i intencions), que és una habilitat indispensable per a les relacions interpersonals. Tanmateix, quan els processos de categorització extrapolen la distinció individual ("jo" i "tu") a una distinció de grup ("nosaltres" versus "ells"), s'esbiaixen les relacions interpersonals, amb una preferència indiscriminada per aquelles persones que pertanyen al grup propi. Tot i que la categorització social és necessària per facilitar als infants la representació del seu entorn social (Dunham i Olson, 2016), sovint desencadena representacions estereotipades dels individus, no atenent a qui són sinó a quin grup pertanyen.

Les respostes dels infants en paradigmes de falses creences (FC) ha estat objecte d'un acalorat debat durant els darrers trenta anys. El fracàs sistemàtic en les tasques estàndard de FC abans del 4t aniversari s'ha interpretat de maneres divergents: els nativistes l'atribueixen a les altes demandes de processament de la tasca en relació amb els recursos executius i d'atenció dels infants d'aquesta edat (Baillargeon et al. , 2010; Leslie, 2005; Leslie et al., 2004; Leslie et al., 2005; Scott et al., 2010); els empiristes atribueixen el fracàs a la manca d'un concepte meta-representacional de la creença (Wellman, 2001; Gopnik i Wellman, 1992) i els pragmàtics atribueixen les respostes incorrectes dels infants als elements contextuals específics de la tasca que la fan pragmàticament difícil pels infants (Helming et al., 2016; Rubio-Fernández & Geurts, 2013; Siegal i Beattie, 1991; Westra, 2017).

L'Outgroup Advantage Pragmatic Account (OAPA) que es presenta en aquesta tesi és una proposta pragmàtica que intenta explicar la dificultat dels infants en la tasca de FC a través dels coneixements de la branca cognitiva de la Teoria de la Ment Intergrupal. L'objectiu de la proposta és mostrar que el rendiment en les tasques estàndard de FB no només està vinculat a la competència dels nens per atribuir creences falses, sinó que està entrellaçat amb factors pragmàtics que fan que la tasca sigui exigent cognitivament. L'OAPA estableix que la pertinença del grup social del receptor de la mentalització (l'individu que té la falsa creença) és un factor pragmàtic que pot influir en el rendiment dels nens en una tasca que enfronta directament la perspectiva del protagonista amb la del participant. Prediu que un protagonista d'un grup extern redueix les exigències pragmàtiques de la tasca de FC perquè presenta un objectiu més fàcil a qui atribuir les falses creences per tres motius: un protagonista del grup extern augmenta la rellevància de la perspectiva del protagonista, augmenta la ignorància esperada del protagonista i aclareix la intenció de l'experimentador.

El treball experimental està dividit en quatre apartats. A la primera secció vaig explorar la influència pragmàtica de manipular la pertinença del grup social del protagonista amb una categoria social real (raça) i una categoria artificial (creada mitjançant el paradigma del grup mínim - Minimal Group Paradigm) en el rendiment en tasques de falses creences d'infants

de 4 a 5 anys. Vaig trobar que un protagonista del grup extern ajuda als participants a obtenir una millor puntuació en la tasca de falses creences anomenada "continguts inesperats" (Gopnik & Astington, 1988) (Estudi 1) i influeix en el patró seqüencial de dificultat de l'escala ToM de cinc ítems de Wellman i Liu (2004) (Estudis 1 i 2).

La segona secció es va dissenyar per abordar el primer factor pragmàtic de la proposta: la rellevància de la perspectiva del protagonista. Els estudis 3 i 4 van confirmar que la manipulació de la rellevància de la perspectiva del protagonista influeix en la capacitat de superar la tasca de "continguts inesperats" d'infants de 4 a 5 anys.
La tercera secció va investigar el segon factor pragmàtic de la proposta: l'expectativa d'ignorància del protagonista. L'estudi 5 va demostrar que la pertinença a un grup compartit entre el protagonista i l'agent fa que la "tasca de canvi d'ubicació" (Wimmer i Perner, 1983) sigui més exigent cognitivament pels infants de 5 a 6 anys.

La quarta i darrera secció tenia com a objectiu examinar la trajectòria de desenvolupament de l'avantatge del grup extern en una tasca de FC contínua amb una població que ja dominava les tasques estàndard de FC. Els estudis amb infants de 8 a 9 anys i adults no van mostrar evidència d'un avantatge de grup extern.

La suma dels resultats dels set estudis presentats en aquesta tesi confirma les prediccions de l'OAPA en el rendiment dels nens de 4 a 6 anys en tasques de FC. L'OAPA dona suport a una visió més flexible del paper que juguen la competència i el rendiment en les tasques de FC i suggereix un camí per desenvolupar estratègies d'intervenció per prevenir els biaixos intergrupals durant la infància.

Table of Contents

List of Abbreviations

Abbreviation	Definition
ANOVA	Analysis of Variance
DB	Diverse Beliefs
DD	Diverse Desires
FB	False Belief
KA	Knowledge Access
IAT	Implicit Association Test
ICC	Items Characteristic Curves
IRBT	Implicit Racial Bias Test
MGP	Minimal Group Paradigm
OAPA	Outgroup Advantage Pragmatic Account
PPVT	Peabody Picture Vocabulary Test
RAE	Real-Apparent Emotion
TB	True Belief
ToM	Theory of Mind

List of Tables

List of Figures

Note: All the illustrations that appear in this dissertation are recreations of the materials that were used for testing. The testing materials were digitally created with Vyond (Vyond, 2017) but the author did not have copyright usage for print. The illustrations that appear in this dissertation have been specifically created to be included in the present work and the author is the owner of the content.

Chapter 1.
General introduction

1.1. Overview

Acquiring the notion of "self" versus "the other" during the early preschool years has a crucial impact on two socio-cognitive processes that have, for the most part, been analyzed separately. One is the development of Theory of Mind (ToM) which is the ability to represent another person's mental states (e.g., beliefs, desires, intentions), and the other is the development of social preferences and category-based social biases.

While understanding there is a "me" and a "you" is key for the development of ToM, which is in turn key for social interactions and a predictor of social ability (Astington & Jenkins, 1995; Davis-Unger & Carlson, 2008; Mizokawa & Koyasu, 2015; Slaughter et al., 2015), understanding there is a "me" and a "you" also comes with dividing the world into "us" vs. "them," which, though necessary to navigate the social world, can also bring along negative consequences in the form of stereotypes and biases that will become deeply rooted and hard to relinquish.

Bridging together the breadth of research from intergroup cognition and ToM development, this dissertation will explore the cognitive implications of the development of intergroup ToM, specifically, it will explore the pragmatic influence that intergroup cognition has on children's performance in false-belief tasks (FB tasks) during development.

The FB task has been established over the last forty years as the litmus test for ToM. Success in this task is determined by the ability to register the difference between one's own experience and the protagonist's experience. The protagonist is the "false-belief holder" or the recipient of the participant's mentalizing projection, and it differs from the participant in that they are ignorant to the situational knowledge the participant has. Although the understanding of the protagonist's "otherness" (distinct from "self") and therefore the protagonist's lack of access to the situational knowledge is indispensable to passing the FB task, few studies have investigated the influence that manipulating the protagonist's group membership in relation to the participant's group membership might have in the mentalization process during development. To date, the "otherness" has been manipulated in terms of the protagonist's age (Seehagen et al., 2018), the protagonist's competence (Zmyj & Seehagen, 2020) and the protagonist supernatural abilities (Lane et al. 2010). The protagonist's group membership in relation to that of the participant has been explored with

the protagonist's cultural and minimal group affiliation (Sudo & Farrar, 2020) and the protagonist's accent (Witt et al., 2022).

Developmental research has shown that social categories emerge quite early during development and often bring along implicit and explicit biases that can lead to a preferential treatment towards the ingroup (e.g. resource allocation, friendship favoritism, competence attribution) as well as a (de)humanization and homogenization of the outgroup (Dunham, 2018a). Considering the weight social categories have during development, it is reasonable to question the influence that the protagonist's social group in relation to that of the participant might have in the mentalization process.

My theoretical approach falls within the realm of pragmatic accounts that aim to alter some pragmatic components in FB tasks to improve children's performance (Helming et al., 2016; Rakoczy & Oktay-Gür, 2020; Rubio-Fernández & Geurts, 2013; Westra & Carruthers, 2017).

Here I am interested in planting a new seed in Developmental Intergroup ToM research by manipulating the protagonist's social group membership to enhance the protagonist's "otherness" and explore whether performance varies under this pragmatic manipulation.

I will first establish the theoretical framework that has founded my research by separately reviewing the most relevant theories and findings in the fields of intergroup cognition and ToM. I will then frame my approach within the field of Developmental Intergroup ToM and will present my research objectives and predictions. The experimental evidence will be presented in the order of study completion and will be examined together in the discussion section. I will conclude with directions for future studies.

1.2. Intergroup cognition

Research on the development of social cognition is intrinsically guided by the study of social groups and the interaction that takes place between and within those groups (Dunham, 2018a). Categorizing people into groups is a core psychological process that develops during infancy and guides how children position themselves in the social system they have been born and exposed to (Liberman et al., 2017). It is a process that continues to be crucial during the entire adult life (Kawakami et al., 2002; Rhodes & Baron, 2019).

Though the notion of group should inherently be bound by its circling entity – you either belong to it or not – this early-developing capacity to assort social interactive partners is an online process that is constantly updating, sometimes leaving the group's entitativity - defined as the group's perceived "groupedness" in Straka et al. (2021)- to the capricious mind of the specific individual doing the sorting. Not surprisingly people can simultaneously belong to multiple groups and can be put in divergent ensembles depending on who they are interacting with or to whom they are being compared to. Even though entitativity research shows that people who are building a house together are perceived as "groupier" than people waiting for the bus together (Plotner et al., 2016), to the mind of the bus driver, those people with angry faces that have been waiting at the stop for a long time, represent a very solid group united by their shared threatening stares.

Moreover, actively categorizing the physical world tends to minimize the variation among the category's individual constituents and puts emphasis on the causal and functional roles that mesh them together. Using the category "chair" as an example, Dunham (2018a), explains that understanding categorization demands acknowledging two complementary questions: a. What features must an object have to be put in a category? b. Once I know that an object falls into a category, which features should it have? The author illustrates the asymmetry of the questions: "we have no trouble identifying a mid-century modern clear plastic wave as a chair but would be unlikely to expect a chair we have not yet seen to possess those idiosyncratic features" (Dunham, 2018a, p. 1).

Similarly, the arbitrary choosing of one feature over another to group individuals together makes the process random and culturally permeable to a fault. We expect intragroup variation to be minimal and focus instead on intergroup differences, carefully stuffing meaning into categories that are initially conceived in an abstract sense and are void of reality.

Ultimately, what we are incessantly doing is dividing the social world between "us" (the ingroup) and "them" (the outgroup). And this would not pose an issue if it were not for the blinding social preferences that categorizing potentially imbues. The construction of the "other" is subsequential to the emergence of the "self," and the mechanisms that humans use to bolster the "selves" can work detrimentally in the regard and assessment of the "others."

1.2.1. Early developing social preferences

It is important to note here that long before the emergence of the "self," infants already show social preferences to similar or familiar others (Kinzler et al., 2007). Though these preferences are not yet determined by conceptually rich social categories (Liberman et al., 2017), they have perceptual and cognitive lasting effects that make the later emerging category-based social preferences even more stringent. Developmental intergroup theory (Bigler & Liben, 2007) postulates that children can express positive attitudes toward a particular group before the conception of that group is fully developed (Dejesus et al., 2017).

It starts from a perceptual standpoint: at 6 months old, infants preferentially look at their racial and linguistic ingroups over outgroups (Kelly et al., 2005; Kinzler et al., 2007, Liu et al., 2015). The preferential looking as well as infants' neural responses may be driven by comfort or liking, and it is influenced by the infants' exposure to ingroup-outgroup exemplars (Hwang et al., 2021). By 9 months, infants start having trouble telling apart racial outgroup faces if they are not exposed to them (Anzures et al., 2013; Quinn et al., 2016), an effect known as "the-other-race effect" that is the result of the perceptual narrowing process that takes place during this first stretch of development. This means that infants' perceptual attunement to their own race's physiognomy brings them to code faces from different races into a broad cluster of "otherness" (Hwang et al., 2021)[1]. Moreover, around the same developmental period, they begin to associate positive valence with their racial ingroups and negative valence with their racial outgroups (Xiao et al., 2018).

Exposing the ontogenetic origins of children's understanding of social categories through perceptual and cognitive research is key to mapping the social preferences that develop with

---

[1] The term "race" is the term used in most intergroup cognition research papers across the world. The term has a pejorative denotation in the country where I am writing this dissertation and I want to clarify to the reader that I will be using the term in an academic manner.

age. Preferences that may begin due to perceptual processes such as processing ease and stability (Freeman & Johnson, 2016) are further amplified by the emergence of the notion of "self" and the child's consequent integration of social categories and assimilation of rich belief systems about those categories. Liberman et al. (2017) sustain that, in the elaborated construction of children's social world, social preferences might lead to prejudice and social categorization might lead to stereotyping. This social construction is nonetheless malleable, as it is highly influenced by exposure, experience, and input (Liberman et al., 2017, Hwang et al., 2021). The level of malleability of these processes is the focus of experimentally designed interventions that aim to reduce ingroup biases.

1.2.2. Self-awareness

Research on the advent of self-awareness and the distinction of the "self" from the "other," supports that an infant reaches this developmental milestone between 18 and 24 months of age (Southgate, 2020).

In her Altercentric Cognition account, Southgate (2020) explains that, before this age, infant cognition is mainly altercentric; the infant's cognitive processing is significantly influenced by the presence of other agents and how these other agents are perceiving and representing the world around them. The author poses that aligning one's attention with the other's attention (usually the mother's) during the early developmental stages may have an adaptative value for guiding infants on selecting what information to acquire. Southgate points to a hypothesis from an evolutionary perspective (Braten, 2002; 2004) that can explain the adaptative value of aligning one's attention with the mother's attention. According to this perspective, altercentrism was the solution selected during hominin evolution to counter the fact that humans do not benefit from spatially sharing their mothers' perspective because, unlike chimps, they can't ride on their mother's back and see "through her eyes" (Southgate, 2020, p.30). Gaze following, gaze cueing, better memory and preference for the targets of other's attention and actions, and motor mimicry are examples of the effects of altercentric perception early in ontogeny (Kampis & Southgate, 2020).

Ultimately, relying on these shared mechanisms blurs the infant's distinction between self and other. It is not until they are two years old that they begin to distinguish between their own experience and the other's experience (Kampis & Southgate, 2020).

One of the main indicators of the emergence of self-awareness is the Mirror Self-Recognition test (MSR) (Gallup, 1970). In this behavioral task the experiment marks the infant's face without them noticing and places the infant in front of a mirror. The premise is that if the infant recognizes the image in the mirror as a reflection of themselves, they will try to touch or remove the mark from their faces. This ability first emerges in infants at the one-year mark and is widespread by the two-year mark (Nielsen et al., 2003). Performance on this task coincides with other abilities related to self-awareness like personal pronoun use and brain indicators of self-awareness (Southgate, 2020).

1.2.3. Us vs. them

The combination of the emergence of the self and the previously acquired familiarity-driven perceptual social preferences, leads to a very strong cocktail for the establishment of ingroup bias. Put simply, an ingroup bias is preferring those that one considers members of one's group. This type of social preference is referred to as a bias because it is disproportionate and usually not founded on objective reasoning.

Interestingly, it applies to groups that could potentially have been socially learned (e.g., social categories such as race, gender, nationality) as well as groups that have been artificially created on the spot and where the only difference between those two groups is that the subject has randomly been put in one and not the other. This effect has been tested through the minimal group paradigm (MGP) (Otten, 2016; Tajfel, 1971) and it shows that "mere membership" is enough to elicit an ingroup bias (Dunham, 2018b).

Why do humans prefer other humans that have been randomly been put in their group? Dunham (2018a) points to two sets of theories that resort to inverse mechanisms to explain the stripped-down ingroup bias. On the one hand, belonging to a group can be a way of bolstering the self: for instance, Social Identity Theory poses that group membership is an approach to positively construe the self (Ellemers & Haslam, 2012). Pertaining to a social group enhances the perception of the self and therefore boosting the qualities of the ingroups brings more worth to the group which then brings more worth to the self (Dunham, 2018a). In this set of theories, the group's high value permeates to the self.

Contrarily, in the other set of theories, it is one's high value that is transferred to the group. Dunham (2018a) alludes to a review by Greenwald and Banaji (1995) in which they gather different examples of individuals' highly positive view of themselves (such as overestimating

their skills or preferring letters in their own names). Self-related positivity, spreads to the group, increasing the ingroups' value. Thus, ingroup favoritism is positively associated with self-esteem (Aberson et al., 2000), and humans continuously make themselves like the people around them (Buttelman, 2013) tracking down positive stories about their ingroup (Over et al., 2018).

Regardless of the direction of the boosting effect, the mere membership ingroup bias develops quite early during development and remains strong even in the absence of any relevant social context during adulthood (Dunham et al., 2011). There is evidence that 3-year-olds already show the MGP effect (Richter et al., 2016). The wide range of ingroup bias effects caused by the MGP that have been reported in the literature can be divided into evaluative effects, coalitional effects and learning effects (Dunham, 2018b). Examples of evaluative effects caused by the MGP are found in studies that show positive explicit attitudes towards an ingroup member (Dunham, 2013) and studies that test friendship preferences (Plötner et al., 2015). Coalitional effects of the MGP have been reported in paradigms that test favoritism on costly giving (Stagnaro et al., 2018) and paradigms that evaluate children's attention to shared goals (Mcclung et al., 2018). Learning effects caused by the MGP have been described in paradigms that assessed memory for bias-consistent information (Dunham, 2011).

It has been extensively demonstrated that the MGP alone can influence behavior and opinions. The perceptual social preferences that appear early in development are enhanced with the emergence of the self and the development of ingroup biases that boost the self. What happens when we add a contextually rich social layer to the mix? That is, what happens during the preschool years with real-world socially constructed categories such as gender or race?

1.2.4. Impact of social categorization processes on social preferences

Social categorization processes are intimately linked with psychological essentialism during development. Psychological essentialism is described as a series of cognitive biases that drive humans to assume an underlying shared essence between members of one group, thinking about them as irrevocably similar (Gelman, 2003; Mandalaywala et al., 2018).

Essentialist beliefs held during infancy shape how children represent and reason about the world around them and the expectations they have for members of the same group. The

following five beliefs constitute the set of essentialist beliefs on categories according to Rhodes and Mandalaywala (2017): a) category boundaries are discrete, b) category boundaries are objective, c) categories mark homogeneous kinds, d) category membership is causally powerful, and e) category membership is intrinsic. These beliefs can have different cultural patterns depending on the salience, meaning and social construction of the social categories (Davoodi et al., 2019) but, generally, they perform as the ultimate guardians of category permanence and immobility. Hence, categories that have been arbitrarily created and are highly dependent on cultural input and historical times become static and limit the social categorization process' much needed flexibility and adaptability.

Since the mid 1990s, essentialist beliefs in early childhood have been measured through "switched at birth" tasks, in which the child needs to decide whether a member of one category (e.g. dog) that is adopted by a member of another category right after birth (e.g. cow) will share the category-typical properties (physical and behavioral) of the birth parents (e.g. dogs) or the adoptive parents (e.g. cows) (Gelman & Wellman, 1991).

In 1995, Hirschfeld adapted the "switched at birth" task to test essentialists beliefs about race (a historically salient social category in the US) and found that 4-year-olds believed that a child would share the physical properties (skin color) of the birth parents, that is a child that was born to Black parents but raised by White parents would look like their biological Black parents. Though this finding has been understood as indisputable proof of children's essentialist beliefs about race, Mandalaywala et al. (2018) point out that there is a difference between acknowledging the heritability of the skin color as something stable (what the child would look like – explored in Hirschfeld, 1995) and the heritability of behavior and personality traits (what the child would act like- not explored in Hirschfeld, 1995). The essentialist beliefs that can influence the development of social attitudes are those of the second kind; the cognitive biases that can drive a child to think that someone's behavior is determined by their race (skin color) and not by their upbringing.

In their research, Mandalaywala et al. (2018) found that while Black and Caucasian 5- to 6-year-olds believe that skin color is inherited (as in Hirschfeld, 1995), they do not hold causal essentialists beliefs about race; they believe that a child's behavioral and psychological properties are inherited by the child's upbringing and not by the child's race. They also found that the beliefs vary by group membership (Caucasians and African-American in their sample) and out-group exposure. The researchers concluded that what prompts an

essentialist belief to become a racial attitude is the interaction with external input from the children's environment and therefore essentialist beliefs can shape different social behaviors for different children (Mandalaywala et al., 2018).

One way in which categories are essentialized through external input is using generic language, which is very frequent in child-directed speech (Cimpian & Markman, 2011). An example of generic language use is when the interlocutor attributes a property to an entire category such as the following statement: "girls like pink." The use of generics creates a tight link between category (e.g. "girls") and property (e.g. "like pink") and reinforces an essentialist view on that category (Gelman et al., 2010). Cimpian & Markman (2011) found that generic assertions are "often accepted on the basis of little to no evidence", "difficult to disconfirm", and "understood properly by preschool-age children" (p.473).

The danger of essentializing social categories during infancy is evident as it provides children with a faulty navigation system that breeds from biases and promotes biases. For instance, continuing with the example of race, the legitimization of this category as a natural distinction between two people through psychological essentialism, makes them assume that existing social hierarchies (that benefit the Caucasian population) reflect a naturally occurring structure (Mandalaywala et al., 2017) that must not be altered.

During the preschool years, children's social preferences are merging with category-based essentialist beliefs about groups, combining stereotypes with prejudices[2], with all the nefarious consequences this process entails. Three-year-old children already show implicit biases based on race (Dunham et al., 2006; 2013; Qian et al., 2016) and these biases remain across development (Baron & Banaji, 2006; Dunham et al., 2013; Gonzalez et al., 2017; Newheiser & Olson, 2012) and into adulthood (Kawakami et al., 2002). Pointedly, prejudice towards other groups increases during development, peaks between the ages of 5 and 7 years, and it progressively declines during middle childhood through the acquisition of flexible thinking (Aboud, 1988).

Keeping a balance between perceptual accuracy, social projection and social categorization is a stalemate conflict during development. While categorical thinking is highly responsible for stereotypical reasoning, it can also be understood as a cognitive heuristic to mentally

---

[2] In social psychology stereotypes are understood as semantic structures that comprise descriptive properties of groups whereas prejudices refer to evaluative structures that encode valence information (Westra, 2019).

represent the world in a less taxing manner (Dunham & Olson, 2016). Furthermore, the mental construal of the "other" is a combination of low-level sensory perception (i.e. processing of facial and vocal cues) and higher-order social cognition such as social categories, stereotypes and high-level cognitive states (Dynamic Interactive Theory of Person Construal, Freeman & Ambady, 2011).

There are multiple studies that focus on intergroup category-based social preferences (both implicit and explicit) and most of them provide evidence for an overwhelming preference for ingroup category members. This applies to social evaluations in different categories such as gender (3-year-olds used gender-based distinctions to choose with whom they wanted to be friends - Shutts et al., 2013), language (5-to-6-year-olds expressed social preferences for native-accented speakers of their language - Kinzler & DeJesus, 2013), race (5-year-olds gave more toys to own-race individuals - Kinzler & Spelke, 2011) and accent (5-to-7-year-olds preferred American-accented English speakers than Korean-accented English speakers - DeJesus et al., 2017).

It is important here to distinguish between a preference or liking for the ingroup member versus a disliking for the outgroup member. Many of the category-based social preferences studies use forced-choice measures which can confound the acceptance of one target with the rejection of the other target (Aboud, 1988). Pun et al. (2017) showed evidence that children's early social group preferences are built from positive evaluations of familiar groups rather than negative evaluations of unfamiliar groups. It seems that ingroup favoritism appears early in development and becomes key to children's intergroup evaluations (Shutts & Kalish, 2021; Aboud, 2003). This tendency remains valid through adulthood as intergroup attitudes seem to be disproportionately influenced by an enhanced evaluation of ingroup members rather than derogation of outgroup members (Mahajan & Wynn, 2012).

Put simply, in the "us" vs. "them" world of intergroup cognition, the "me" in the "us" is the indisputable king. In the words of Rhodes and Baron (2019), "the process of self-identification actually fosters social categorization."

1.2.5. Impact of social categorization processes on behavioral patterns and cognitive skills

As it would be expected, social categories not only influence children's social preferences, but they can get interwoven in patterns of behavior and cognition (Skinner & Meltzof, 2019).

Behavioral patterns

Learning, imitation, sharing and helping are some of the behavioral patterns that may be biased by social categorization processes. These behaviors are built on what has been considered the root of human sociality, which is the engagement in interactions of shared intentionality (Tomasello, 2020).

In this line, Han Li and Koenig (2022) propose that group membership cues serve as both epistemic (expecting relevant and reliable information from ingroup members) and social cues (willingness to collaborate and affiliate with ingroup members). Regarding epistemic cues, it has been demonstrated that children selectively learn from their ingroup members, and that they do so even when they are shown antisocial behavior of ingroup members (Mills & Elashi, 2014; Hetherington et al., 2014; Kinzler et al., 2011; Wilks et al., 2019).

As for collaboration, it has also been observed that very early during development, infants start to show cooperative behaviors in situations that require joint perspectives and shared goals (Tomasello, 2020). Many studies have revealed that the parsing of the social world has an influence in these cooperative behaviors: children expect a) ingroup members to cooperate and help one another more than outgroup members (Rhodes & Baron, 2019), b) agents to share resources with ingroup members (Bian et al., 2018), c) group members to support one another in intergroup conflict (Chalik & Rhodes, 2014), d) they are not willing to help agents that have previously harmed ingroup members (Ting et al., 2019) and e) they refer to group membership as a reason to inflict harm on an outgroup member and help an ingroup member (Chalik & Rhodes, 2015).

Other behavioral consequences of social categorization are related to performance and academic accomplishments (Rhodes & Baron, 2019). A clear example is math: holding gendered stereotypes on math performance (i.e., the widely extended "boys are better at it than girls") predicts girl's lower math self-concept and math achievement during the elementary school years (Cvencek et al., 2015). In the same line, six-year-old boys are more likely than girls to believe their own-gender group members to be "brilliant" which leads girls to start dismissing activities that are for children that are "really smart" from a very young age (Bian et al., 2017).

Cognitive skills

The example of "deception"

The cognitive ability to deceive others (instilling a false belief in another person) emerges early during development and it improves with age, specifically between the ages of 3 and 7 years old (Lee, 2013; Pan et al., 2015). This ability is considered both an adaptive and maladaptive skill that can be used both for prosocial purposes and antisocial purposes (Lee, 2013). Interested in observing such a complex social cognitive skill through an intergroup lens, De la Cerda and Warnell (2020) examined whether the identity of the social partner (the lie-receiver) influenced children's ability and willingness to lie. It was found that children's lying skills were flexible based on group membership and contextual demands: they were equally likely to lie to ingroup and outgroup members in a self-benefit scenario. However, they were more likely to lie to an outgroup member in a scenario where someone else benefited from their lie and, more importantly, in a scenario where there was no gain, either for self or for the other (De la Cerda & Warnell, 2020). That is, lying to an outgroup member came more naturally than lying to an ingroup one.

Children's lie-telling abilities have been long linked to their executive functioning development (Talwar & Crossman, 2012) and it has been proven that Theory of Mind (ToM) understanding has a direct influence in the development of children's deception capabilities (Ding et al., 2015; Talwar & Crossman, 2012; Talwar & Lee, 2008; Talwar et al., 2007).

A crucial question arises from all the previous studies and results: how do social categorization processes influence the development of ToM understanding? In the next chapter I will briefly review the extensive research on ToM development to set the stage for bringing the two disciplines together from a theoretical point of view.

Box 1: Key points from Intergroup Cognition.

- Categorizing people into groups and diving the world between "us" vs. "them" is a core psychological process that develops during infancy. Even before the emergence of the self.

- Early developing social preferences are perceptually based and highly dependent on exposure, experience, and input.

- Mere membership is enough to elicit an ingroup bias.

- Essentialist beliefs shape how children represent the world around them and the expectations they have for group members.

- Assorting social interactive partners is an online process that is constantly updating.

- Humans expect intragroup variation to be minimal and focus on intergroup differences.

- During the preschool years, children's early social preferences interact with category-based essentialist beliefs about groups, resulting in stereotypes.

- Prejudices peak between 5 and 7 years.

- Learning, imitation, sharing and helping are some of the behavioral patterns that are biased by social categorization processes.

- Lying is a cognitive skill affected by group membership.

1.3. Theory of Mind

1.3.1. Origins and role in social cognition

The term Theory of Mind (ToM) was first coined by Premack and Woodruff in 1978 to explain children's understanding of intentional behavior, a research field that Piaget had started investigating a few years prior, in 1973 (Glenn & Parry, 1993). By ToM they referred to the ability to impute mental states to oneself and to others through a system of inferences. They included the term *theory* because, on the one hand, mental states are not directly observable and, on the other hand, such system can be used to make behavioral predictions (Glenn & Parry, 1993). There are multiple mental states that have been gathered under the ToM umbrella (e.g., beliefs, desires, knowledge, intentions, and emotions) and, during the last forty years, developmental researchers have dedicated significant efforts to unveil the intricacies of when and how children develop the ability to attribute these mental states to others.

Steady interest in this research domain is due to the weight ToM has in the development of social cognition and social relationships starting in infancy and throughout one's life (Karmakar & Dogra, 2019). Lecce and Devine (2022) point to two main reasons why this field has drawn so much interest from the research community: one is the variation within ToM capabilities that can be a guide to the child's social cognitive development; ToM performance is correlated with perceived social competence by teachers (Devine et al., 2016), with peer acceptance (Banerjee et al., 2011) and reciprocated friendships (Fink et al., 2015). The second one is ToM's malleability; the fact that ToM can be influenced by social experience (Devine & Hughes, 2018) and can be improved through training interventions (Lecce et al., 2014).

Moreover, ToM is also a key indicator of the adequate development of domain-general cognitive skills. Passing a standard ToM task directly involves three very important executive functions: a) it requires working memory to be able to hold and process the information about the characters, and it draws from 2) inhibitory and 3) attentional control to suppress the child's own accurate perspective and attend to the target's mistaken perspective (Grace & Kim, 2020).

Reasoning about the thoughts of others is indispensable to cooperate and form complex social systems (Hoyos et al., 2020). Interestingly, ToM proficiency operates simultaneously

at two seemingly contradictory levels: one the one hand a good ToM command boosts social aptness (e.g., it is associated with popularity in early childhood - Slaughter et al., 2015 - as it facilitates effective prosocial interaction - Spenser & Winder, 2020 -); on the other hand, it improves children's ability to lie (Ding et al., 2015). Thus, it can be a key ingredient for empathy (empathic understanding being comprised of cognitive empathic understanding (ToM) and affective empathic understanding - necessary to allow and individual to share another person's emotional experiences - (Spenser et al., 2020), as well as a key ingredient for deception and manipulation.

Hence, ToM is a set of cognitive abilities indispensable for interpersonal relationship building and strategizing that needs to be efficient and flexible as it operates simultaneously in both *competitive* and *cooperative* settings (Apperly & Butterfill, 2009).

If humanity has developed ToM skills to compete and cooperate, it would be interesting to explore whether these skills are enhanced when interacting with ingroup individuals and/or outgroup individuals.

1.3.2. The False Belief Task (FB task)

The prevalent experimental designs for ToM development during infancy are the verbal FB tasks. The first FB task was created by Wimmer and Perner (1983) and it is also known as the change-of-location FB task. One of the most dominant versions of the task is called the Sally-Anne test and it was created by Baron-Cohen et al. (1985) only two years after its conception (see Figure 1).

The task goes as follows: at the beginning the participant is presented with two agents (they can be puppets or cartoons); Sally and Anne. Sally has a container and Anne has another container. Sally places a marble in her container and leaves the scene. While Sally is away, Anne moves the marble from one container to the other. When Sally comes back, she wants to play with her marble. And the question the participant is asked is: "Where will Sally look for her marble?" To pass the task, the participant needs to ignore the information they now have about the location of the marble and be able to represent what Sally thinks or will do.
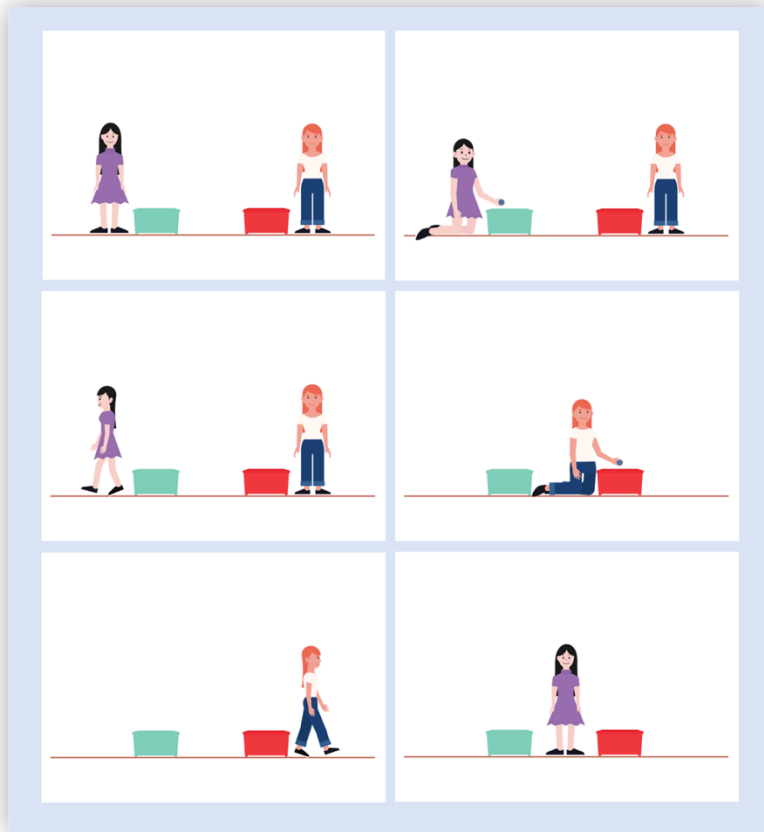
Figure 1. Illustration of the Sally-Anne task.

The second FB task that has become a standard test to assess false-belief understanding is called the unexpected-contents task and it was first conducted by Gopnik and Astington (1988). It is also known as the "smarties" task (see Figure 2). In this task the participant is presented a box of candies (smarties) and is asked what they believe is in the box. After the child (usually) guesses that there are smarties in the box the experimenter opens the box and shows the participant that the box contains pencils. The experimenter then closes the box and asks the child what they think "someone" (i.e., the mentalizing target, someone who is not present and has not seen the content of the box) will think is in the box. To succeed, the participant needs to ignore the very salient situational information they now have and respond as they initially responded, that there are smarties in the smarties box.
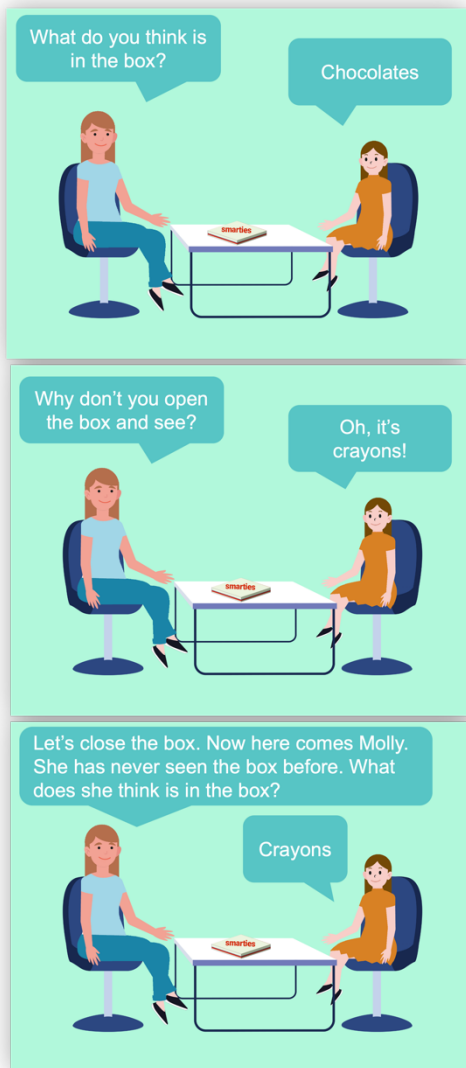
Figure 2. Illustration of the "smarties" task.

There have been multiple iterations of these two tasks over the past 30 years and the wealth of data generated shows that children begin to consistently pass the FB tasks around 4 years of age (Wellman et al., 2001). Before that age, children usually respond with the contextual information they have, that is: 1) Sally will look for her marble where the participant has last seen it 2) the target thinks that there are *smarties* in the band-aid box (Wellman et al., 2001). They are called FB tasks because to give a correct answer, children cannot assess reality and use that information to infer the other's knowledge, instead children need to understand that a person can hold a belief that is not in line with reality, that is, that the "other" can hold a different belief that oneself, a "false belief". Being able to pass the task has been considered evidence for distinguishing between mind and world (Wellman et al., 2001), understanding

that people can act not according to a situation but rather according to the mental representation of that situation (Glenn & Parry, 1993).

Some studies have shown that the robustness of these tests remains airtight to modifications on the way it is presented. Wellman et al. (2001) conducted a meta-analysis of 178 studies that had used FB paradigms and found there was no significant difference in performance between the two task types (change-of-location and unexpected-contents). Likewise, neither the medium in which the characters in the story were presented (e.g., puppets, dolls, cartoons, or real persons) nor the way in which the object was depicted (e.g., as a picture, in a video or as a real item) altered performance. Further, they also showed that children's responses appeared to be consistent across trials and their overall performance was not altered by the phrasing of the question, from using the expressions "will look for", "will think" to pointing with the finger or even providing a yes/no answer (Wellman et al., 2001). They concluded that performance in the FB tasks showed an indisputable developmental pattern; preschoolers went from a below-chance performance (at 3 years) to an above-chance performance (at 4 years). They conceded that performance at 3-years-old could be improved with some task manipulations, but it could not be lifted to above chance levels.

Nonetheless, other recent studies have shown that children's response to the FB task is not entirely consistent through consecutive trials, incurring in wide variability in performance and dataset noisiness, suggesting that children may often have inconsistent or partial mastery of ToM concepts (Richardson et al., 2018; Warnell & Redcay, 2019). Moreover, modified versions of the change-of-location FB task brought 3-year-olds performance to above chance levels (Rubio-Fernández & Geurts, 2013; 2016).

1.3.3. Performance in the verbal FB tasks: interpretation of success

Despite the extended use of these tasks, it is important to note that ToM should be considered as a broader system of inferences, and the FB tasks only focus on a specific component of ToM which is false-belief representation (Karmakar & Dogra, 2019). Furthermore, even when the FB tasks are examined solely as an assessment of false-belief representation, there is not a consensus on whether passing the FB tasks marks the acquisition of FB understanding or if such ability emerges earlier but may not be captured by the paradigm.

The systematic failure of FB tasks before the 4th birthday is indisputable. What is up for debate and has divided psychologists and philosophers along empiricist and nativist paths is how to interpret a) the shift in task performance during development (Westra, 2017) and b) variability in performance once they start passing the task (Apperly, 2012). Apperly (2012) explains that a very common pattern of results in two consecutive FB tasks from children that are within a sensitive age range (3-to-6-year-olds) can be a score of 2/2, 1/2 and 0/2. Notably, each school of thought interprets variability through a different lens.

The empiricist approach

The predominant interpretation for the FB task's performance from its first research run up until the early 2000s were empiricist accounts. Under the name of the conceptual-change account (Wellman, 2001) or the cultural constructivism account (Heyes & Frith, 2014), in general terms, empiricists argued that the performance shift in FB tasks at the age of four marked the acquisition of the meta-representational concept of belief (Gopnik & Wellman, 1992; Westra, 2017a). Before that age, children are not equipped to hold a meta-representation of belief.

Passing the task and thus acquiring the ability to predict another person's actions based on that person's mistaken belief is interpreted as a conceptual leap, the obtainment of a new cognitive mechanism that did not previously exist (Richardson et al., 2018), that is culturally learned through verbal interactions, therefore linked to language acquisition (Helming et al., 2016).

For empiricists/conceptualists (e.g., Wellman et al. 2001; Wellman & Liu, 2004; Gopnik & Meltzoff, 1997), 2/2 scorers are assumed to have a conceptual understanding of belief and 0/2 are assumed to not have a conceptual understanding of belief. Children that score 1/2 are difficult to interpret as conceptually it is impossible to have and not have the concept of belief simultaneously. For this reason, such results are attributed to measurement errors (Apperly, 2012).

The nativist approach

On the other side, nativists attribute children's failure to pass the FB task before 4-years-old to the high processing demands of the task and children's underdeveloped executive and attentional resources before that age (e.g., Baillargeon et al., 2010; Leslie et al., 2004; Leslie et al., 2005; Scott et al., 2015; reviewed in Westra, 2017a). The intrinsic demands of the task require sophisticated verbal and executive-functioning skills and mask children's actual competence in FB understanding, resulting in poor performance.

Also known as the early-competence accounts, or the processing-load accounts (Baillargeon et al., 2010), the evidence provided to support their nativist position consists of using violation-of-expectation and anticipatory looking paradigms in which preverbal infants expect others to behave in line with the contents of their false beliefs, showing that false-belief understanding exists before the 4-year-old mark. (Baillargeon et al., 2010; Onishi & Baillargeon, 2005; Scott & Baillargeon, 2017).

For nativists, since inhibitory control is crucial for passing the task, children who score 2/2 are assumed to have the necessary inhibitory resources, while those who fail both trials are assumed to lack the sufficient resources; the assumption for the 1/2 scorers is that their inhibitory control is developing but has not entirely matured yet (Apperly, 2012).

The Two-Systems Account

The Two-Systems Account (Apperly & Butterfill, 2009) emerged from the need to find a middle ground between the evidence put forward by the empiricist and the nativist accounts. The authors proposed a two-systems account for mind reading composed by a) an innate system that is cognitively efficient but inflexible and limited to simple problems (that would explain infants' success in non-verbal anticipatory looking FB tasks) and b) a system that develops over time that is more flexible but more cognitively demanding as it is influenced by demands on attention and working memory as well as language development. Whereas System 1 processes "belief-like" states and it is the system that is recruited in anticipatory looking paradigms in infancy, System 2 deals with actual beliefs and is the system that needs to be developed to pass FB tasks.

Beyond the Two-Systems Account, there is another approach that scientists have taken to explain the systematic failure of FB tasks before children's fourth birthday that also offers a more flexible explanation for the ½ scorers: the pragmatic accounts.

The pragmatics accounts

Wellman et al. (2001) explain that performance on a cognitive task reflects two factors: competence and performance. In their words, competence is the "conceptual understanding required to solve the problem" and performance is comprised by "all the other non-focal cognitive skills (e.g., ability to remember the key information, focus attention, comprehend, and answer various questions) required to access and express understanding" (p.657).

On one side, the nativist accounts focus on showing that competence is already there before the 4-year-old mark by providing research data using alternative paradigms catered to the youngster's cognitive abilities and based on non-linguistic behaviors. On the other side, the empiricist accounts focus on finding consistent patterns and scalable tasks that are passed gradually along with children's conceptual development of meta-representation.

Pragmatics accounts focus instead on analyzing the specific contextual elements of the task (e.g., pre-existing knowledge, inferred intent of the experimenter, participant's implication) as well as the linguistic elements of the question (e.g., the use of the verb "think") that might temper with the efficiency and accuracy of children's information-processing strategies. Put simpler, they focus their attention on why children fail FB tasks from a pragmatic point of view. Some of the pragmatic accounts are based or ultimately lean towards the empiricist or nativist side but others transcend the debate and focus on the pragmatic issues of the task. Though often theoretical, the pragmatic accounts can suggest or present evidence of task manipulations that can boost or worsen task performance.

Conversational implicatures, a Gricean account

Siegal and Beattie (1991) were the first to attend to the pragmatics of the question to explain children's difficulty in responding. They based their hypothesis on Grice's maxims of quantity "Speak no more or less than is required" and relation "Be relevant" in conversation (Grice, 1975). The authors propose that the conversational implicatures of the test question are the cause of children's failure to pass the task. For children to succeed in the change-of-location FB task, they need to understand the purpose of the experimenter's question which is to test them on a mentalistic projection of a deceived target that could be more specifically

framed as "Where will Sally look for her marble <u>first</u>". Instead, children might be understanding the test question in a more world-oriented manner - whether they can predict the behavior of the target in achieving a goal - which leads them to respond to the complete question "Where will Sally have to look for the marble to find it." Their task manipulation layering in the temporal aspect and including "first" at the end of the test question brought the 3-and 4-year-olds' performance to above chance levels: 35% passed the regular task and 71% passed the new version of the task (Siegal & Beattie, 1991).

The Pragmatic Development Account

This theoretical account provides a revised nativist proposal for ToM development through a pragmatics lens. Westra (2017a) proposes that children do not lack a concept of belief before they start passing the FB task but rather, they are not "good at understanding when facts about belief are relevant to conversation" (p.237). Children can answer the question because they are capable of tracking and maintaining the agent's beliefs, but they fail to understand that "the experimenter is interested in those beliefs" (p.246). Building off from Siegal and Beattie's (1991) research, the author suggests there are three interpretations of the test question in a change-of-location FB task and only one of them provides a successful answer. The first one is that the child is being asked to be helpful to the protagonist, the second is that the child is being asked to exhibit their knowledge of the events that have been narrated, and finally the third one is the intended interpretation about exhibiting their knowledge on the protagonist's beliefs.

Pragmatic Performance Limitation Account

The Pragmatic Performance Limitation Account (Rakoczy & Oktay-Gür, 2020) aims to explain the surprising finding that when children begin to pass the FB task, they start to fail the True-belief task (TB task) which they mastered until then. The TB task follows an exact set-up to the FB task but in this version the protagonist stays in the scene while the agent moves the object, which makes the answer to the test obvious. The authors sustain that there are pragmatic factors that make the test question of the TB task very odd, and they try to explain the U-shaped developmental pattern of performance curve through these factors. The triviality of the question and the academic nature of the question are among the factors that make the task pragmatically confusing for children between 4 and 7 years who have developed basic meta-representational capabilities but have not acquired the pragmatic

sensitivity necessary to understand all sorts of speech acts. The authors present experimental evidence of improved performance in versions of the TB task where the pragmatic oddities have been removed (i.e., nonverbal versions of the task that eliminate the academic/trivial question or verbal versions in which the triviality of the question is made explicit - "I'll ask you a baby question").

The Pragmatic Framework

Helming et al.'s (2016) pragmatic framework discusses four discrete possibilities that the human mindreading system offers when faced with a change-of-location FB task. First, they establish the distinction between mind-reading the psychological states of agents of instrumental actions (the protagonist of the FB task) versus communicative actions (the experimenter's question). Then, they point to the human ability to take a "detached third-person perspective" or an "engaged second-person perspective" on both instrumental and communicative actions.

Providing a correct response to a change-of-location FB task requires that the participant takes a third-person perspective onto the protagonist's instrumental action and tracks the content of the protagonist's motivation and epistemic state "without intending to contribute to either the success or the failure of the protagonist's instrumental action" (Helming et al., 2016, p.448).

The confusion comes when the experimenter engages in a communicative action with the participant. On the one hand, by asking a question the experimenter emphasizes the epistemic perspective that it is now shared in the room and diminishes focus on the protagonist's mistaken epistemic perspective. The authors refer to this event as "the referential bias" - also called the "shared-information bias" in Salter and Breheny (2019).

On the other hand, the experimenter's question triggers the participant to take a second-person perspective to respond, and the authors suggest that the participant has then trouble to simultaneously stick to the third-person perspective when evaluating what the mistaken protagonist will do. Specifically, taking a second-person perspective onto the protagonist's instrumental action implies that the success or failure of the protagonist's action depends on one own's contribution – which can come in the shape of cooperation or competition.

Helming et al. (2016) point out that very early during development humans are altruistically inclined to take a second-person perspective on other's instrumental actions and help agents

achieve their goals (Tomasello, 2009). Therefore, the reason why young children may be failing the change-of-location FB tasks is that they are cooperating with the protagonist by contributing to their success in achieving their goal (i.e., find the misplaced object). Put simply, they are giving the answer that compensates for the deceived state of the protagonist, they are helping the protagonist find the object.

Hence, the authors hypothesize that young children's performance is compromised by a "cooperative bias" that motivates them to help the mistaken protagonist when they understand that their instrumental action is being compromised by a false belief that has been caused by a second agent.

According to them, the cooperative bias turns the experimenter's prediction question "Where will Sally look for her marble'" to the normative question "Where should Sally look for her marble?" Answering the question normatively means failing to pass the task. Helming et. al (2016) further explain that, aside from the cooperative bias, there is yet another contextual pragmatic element that might turn the prediction question into a normative question. Since the experimenter knows the same information as the child does, they might understand they are being tested on their ability to tell where the mistaken protagonist *should* look for. This hypothesis is also shared by Ben-Yami et al.'s (2019) pragmatic account called "the misunderstanding hypothesis." They pose that the pragmatics of the question incline children to understand the question normatively instead of descriptively, which in an educational setting would lead to "what is the right thing to do."

The Duplo Task: an experimental approach

Rubio-Fernández and Geurts (2013) proposed that children's failure at the standard change-of-location FB task is caused by the test question that disrupts children's perspective-tracking process. The "Where will Sally look for her marble?" draws children's attention to the target object and increases the saliency of the wrong response. They argued that children would be able to track the mentalizing target's perspective if they could focus on them throughout the narrative of the task. To test their hypothesis, they created the Duplo task. They used physical props (Duplo figures) and made two manipulations that reduced the pragmatic demands of the test by helping the participant stay focused on the protagonist's perspective. First, they kept the figure of the protagonist present throughout the task instead of making it disappear as it would usually be done in a regular task. In their manipulation the protagonist turns the back on the scene but stays there. Second, instead of asking the direct standard FB question,

the experimenter placed the protagonist in front of the containers and encouraged the participant to perform an action prompted by the following questions: "What happens next? What is the girl going to do now?" Three-year-old children were able to pass this version of the task.

1.3.4. Performance in the verbal FB tasks: interpersonal cognitive biases

Aside from the biases generated from strictly pragmatic factors, there are other cognitive biases that are triggered in FB tasks that have been studied by researchers from all three schools of thought (mostly nativism). These are interpersonal cognitive biases, which are generated by the inherent presence of a protagonist in a FB task, and they look at the influence that the protagonist's presence imbues in the participant's mental state reasoning.

These biases are crucial to develop the field of intergroup ToM and are the last foundational ground to the theoretical framework of my research. In general terms, the lens through which to read the following biases would be: what happens to these biases when the protagonist belongs to a different social group?

The curse of knowledge
Birch and Bloom (2004, 2007) explain that mental state reasoning is biased from childhood to adulthood by what they call: the "curse of knowledge." They define it as a "tendency to be biased by one's own knowledge when attempting to appreciate a more naïve or uninformed perspective" (Birch & Bloom, 2004, p.255). Such bias diminishes with age and is a major contributor to the fail-success 4-year-old threshold in FB Tasks.

According to them, the curse of knowledge gives two explanations on children's limitations when reasoning about mental states. First, it is a partial bias and not a pure one.[3] For this reason, measures that are categorical (like the standard FB tasks) are not sensitive enough to capture the nuance in children's performance. Second, they point to an asymmetry in knowledge assessment, establishing that it is harder to inhibit one's knowledge (as we tend to think "others know what we know") than to inhibit one's ignorance (as we seldom think

---

[3] The authors provide an example in adult mental reasoning: "Adults who know the earnings of a company, for instance, will show a bias towards the answer that they know when assessing the knowledge of an ignorant person; they do not assume that other people will guess the exact right answer." (Birch & Bloom, 2004, p.256)

"others don't know what we don't know"). In a FB task, the child is aware of the question's answer and for that reason it is harder for them to ignore it.

The altercentric bias

Southgate's Altercentric Cognition account (presented in the first section to introduce the concept of "self" and "other") is also relevant when considering interpersonal cognitive biases. In Southgate (2020), the author proposes the "altercentric bias" as the reason why infants succeed in taking the other person's perspective in non-linguistic FB paradigms. This bias emerges as a combination of two factors: the high value of other's attention during infancy and the absence of a competing self-perspective.

Before the development of cognitive self-awareness that facilitates the emergence of the self-perspective, infants can take others' perspectives without needing the sophisticated inhibitory resources to ignore their own perspective as their own perspective is encoded weaklier than the other's perspective. This is called "the altercentric advantage" and in a FB task this mechanism relies on an enhancement of what the protagonist sees and a strengthened memory for that location instead of the real-world location.

Egocentric biases

Flavell (1977) explains that there is a difference in the way one experiences their own point of view and the other's point of view. While we experience our own perspective in a direct way, we access the other's view indirectly. Our point of view becomes a filter to the other's view as we "are usually unable to turn our own viewpoint off completely when trying to infer the other's" (Flavell, 1977, p.124 – cited in Nickerson, 1999). This noisy filter results in what the author calls "egocentric distortion."

Nickerson's (1999) knowledge model proposal goes a step further. In his view, making sense of another person's knowledge always starts from first establishing a model of one's knowledge and imputing that knowledge to the other person as a default mechanism. The process grows richer when one evaluates the information they have about the other person (categorical affiliation for instance) and establish how it can digress from their default model. This process is egocentrically biased as one will overestimate the other persons' knowledge (closer to their own knowledge) and will underestimate the other person's having more

knowledge than one's own (like the knowledge-ignorance asymmetry in Birch and Bloom's "curse of knowledge").

In line with Nickerson's account, Robbins and Krueger (2005) introduce the notion of "social projection" as a process through which people expect others to be like them. In an intragroup setting, projection can serve as an "egocentric heuristic for inductive reasoning" (Robbins & Krueger, 2005, p.32) and can result in favorable and nuanced perception. In an intergroup setting instead, social projection competes with social category-driven perception that simplifies the process but bundles the other group in a homogeneous group.

Similarly, Ames (2004) proposed a double inferential strategic mechanism for mind-reading: projection (from the self-template) and stereotyping. The author found that the perceived similarity with a target influenced adult's inferences about the target's mental states. An ingroup target activated higher levels of projection of the participant's own mental states versus an outgroup target who activated stereotyping processes instead.

In sum, though projection is generally preferred in terms of valuing and respecting the other and appreciating the human being in whole instead of simplifying the mentalizing effort by applying stiff stereotyping templates, in a FB setting, imputing the self-template to the other can contribute to egocentric blindness and task failure. How do stereotypes interact with mindreading?

Stereotyping and mind-reading

Westra's (2019) theoretical account is possibly the best laying ground from which to build any experimental prediction or theory in the incipient field of intergroup ToM. From an intricate philosophical approach, the author tries to bring together research from developmental psychology and social psychology to analyze the relationship between mindreading and stereotyping as two social cognitive processes that depend on one another and occur simultaneously.

In the words of Amodio (2014), stereotypes are "stored bodies of rapidly accessible semantic information about the generic characteristics and attributes of social groups" (cited in Westra, 2019, p. 2824). As seen in the first chapter of the introduction, stereotypes are already present during the first years of life and become easily accessible shortcuts to understand and parse the world around us.

Westra explains that stereotypes are structured around character traits. Character traits are stable psychological properties that affect behavior in a consistent manner across diverging situations (e.g., laziness, intelligence, honesty) and, according to the Stereotype Content Model, are divided into two main dimensions; the warmth dimension (with traits like friendly/unfriendly or kind/unkind), and the competence dimension (with traits like intelligent/unintelligent or skillful/clumsy (Cuddy et al. 2009).

According to Westra's account, in an intergroup context, projecting an essentialized stereotype to an individual implies attributing a set of predetermined character traits linked to a social group membership. These traits inform inferences about the individual's beliefs and desires which, consecutively, update inferences about their behavior. To exemplify it, an individual that is stereotyped as high warmth will appear to be more helpful and honest without needing a prior social exchange to base this assumption. In Westra's words, "stereotypes save computational resources by providing us with a heuristic for rapidly generating relevant mentalistic interpretations of behavior" (Westra, 2019, p. 2836).

When it comes to the mentalization effort to solve FB tasks, projecting high-competence traits might make true-belief attributions more probable to the protagonist and, likewise, projecting low-competence traits might trigger a prediction of the protagonist's false-beliefs and ignorance. Attributing mental states to protagonists of known social groups then becomes highly dependent on the essentialized stereotypes that one has on these groups. While the saliency of these essentialized stereotypes will depend on the child's experience with other groups and with the child's age (it will diminish as they grow older), mentalizing processes seem to remain similar across groups and stages.

Box 1: Key points from Theory of Mind research

- Theory of Mind (ToM) is a key indicator of adequate development of domain-general cognitive skills.

- ToM is correlated with pro-social behavior and social competence.

- ToM can be used in interpersonal relationships in both competitive and cooperative settings.

- The false-belief tasks (change-of-location and unexpected-contents) are the standard test for ToM during early childhood.

- Empiricists propose that passing the FB task marks children's acquisition of the meta-representational concept of belief.

- Nativists argue that infants can ascribe false belief very early during development and failure at FB tasks is due to the task's high processing demand in executive function skills.

- Pragmatic accounts analyze the pragmatic task factors that worsen performance (e.g. pre-existing knowledge, experimenters' intention, experimenter's question).

- Children's FB task failure might be driven by a cooperative bias that prompts them to help the mistaken agent by providing the real-world correct answer.

- Egocentric biases can also account for children's difficulty with the FB tasks.

- Infants have an altercentric bias in perspective-taking before they develop the notion of self.

- Highlighting the difference between one's perspective and the other's perspective might make the task easier.

- Stereotypes can work as a heuristic for mentalizing.

Chapter 2.
Scope of the dissertation

2.1. Developmental Intergroup ToM

Developmental Intergroup ToM is a slowly growing field that encompasses two kinds of experimental research. Considering how ToM skills correlate with pro-social behavior and social aptness, it is logical that one of the branches of Developmental Intergroup ToM is concerned with the *social* implications that ToM abilities bring to intergroup relationships. In this line of research, Glidden et al. (2021) used the term to circumscribe their analysis of the mediating relationship between mental state understanding and children's ingroup biases. Abrams et al. (2014) found that 6-to-7-year-olds' second-order ToM skills were correlated with their ability to examine the peer status dynamic of a social situation (e.g., deviance from the group and peer acceptance). Mulvey et al. (2016) found that higher mental state skills predict defiance to gender-stereotypic norms in 3-to-6-year-olds, and Mulvey et al. (2021) revealed that children and adolescents with more advanced ToM abilities were more likely to show cognitive forgiveness as well as behavioral intention to forgive towards transgressors than those with less advanced ToM skills. Lastly, Rizzo and Killen (2018) showed that an artificially created social status influenced 3-to-7-year-olds' ToM abilities; children that were assigned to a "disadvantaged" social status condition were more likely to pass the FB tasks.

The other branch of the field is focused on the *cognitive* implications that derive from mentalizing about someone who belongs to one's group membership and someone who does not. Though the research is still quite scarce there have been a few developmental studies that have looked at the difference in children's mentalizing performance when presented with an ingroup or an outgroup protagonist. This question has been addressed using different tasks that tap into different mentalization processes. The performance differential factor between conditions with an ingroup versus an outgroup target seems to be whether the paradigm used taps into a mentalization effort that involves conflicting perspectives between the participant and the target or not.

In paradigms where there is no conflict between perspectives there seems to be an advantage when mentalizing about an ingroup target. McLoughlin and Over (2017) tested the target's membership influence in children's mental state attribution performance with the Frith-Happé animations task (Abell et al., 2000). This task shows a series of animations in which two geometric shapes (a large red triangle and a small blue triangle) move around the screen and interact. It evaluates children's descriptions of the shapes' actions to assess mental state attribution abilities; an action can be described as a physical action (e.g., one triangle pushing

the other triangle) or it can be described using mental state terms (e.g., one triangle is teasing the other triangle). McLoughlin and Over's (2017) manipulated the group membership of the triangles by giving them an ingroup or an outgroup identity along two social categories: geography and gender (i.e., they presented the geometric shapes as "these are two girls/two boys" or "these two children live in the same town as you/different town"). The researchers found that 5-to-6-year-olds spontaneously used more mental state words when describing ingroup members than outgroup members and 6-year-olds produced a larger variety of mental-state terms when talking about their social group.

Gönültaş et al. (2019) found a similar effect testing 9-to-13-year-olds with the Strange Stories task (White et al. 2009). This task presents a series of mental state scenarios that include misunderstandings, jokes, white lies and double bluffs and the participant is tested on whether they show an understanding of how the protagonist would react to a certain scenario (i.e., participants are asked to make a causal inference of the protagonist's behavior). Gönültaş et al. (2019) manipulated the protagonists' group membership to either belong to the participant's cultural ingroup (Turkish) or to an outgroup (Northern European and Syrian). They showed that children exhibited more accurate mindreading for ingroup targets compared to outgroup targets. Moreover, their accuracy in inferring the target's mental states were negatively related to prejudice and perceived threat.

The above evidence is coherent with the theoretical framework revised in the previous chapter; perceived similarity can lead to a more accurate mentalization effort as the participant can derive the mentalization content from their own experience. A more accurate understanding of the ingroup target's mental states could also be enhanced from children's better appreciation of an ingroup target as an individual that is perceived as more human or that sustains human emotions like them. Using an emotion-attribution measure, Costello and Hodson (2014) reported that 6-to-10-year-olds attributed fewer uniquely human emotions to a racial outgroup target. Similarly, in an emotion-recognition paradigm, Segal et al. (2019) found that 6-to-10-year-olds South Asian participants were better at recognizing emotion in own-race faces compared to other-races faces. Lastly, McLoughlin, Tipper and Over (2018) asked 5-to-6-year-olds to rate the humanness of a set of ambiguous doll-human face morphs and found that they perceived racial outgroup faces as less human relative to ingroup faces.

Contrastingly, in paradigms where the protagonist's perspective conflicts with the participant's perspective (i.e., FB paradigms), the influence of group membership in performance has yielded inconclusive results. Studies have shown that the age and the competence of the target influences performance in FB paradigms: Seehagen et al. (2018) found that 4-year-olds were less likely to attribute false beliefs to adults than to peer protagonists[4]; Zmyj and Seehagen (2020) reported that 4-to-5-year-olds were also less likely to attribute false beliefs to competent targets versus incompetent targets. The nature of the mentalizing mind (i.e., humans versus extraordinary beings) has also been shown to influence performance with 5-year-old children as they were less likely to attribute false beliefs to special extraordinary beings than to humans (Lane et al. 2010).

One potential common thread between these studies is that a target that is perceived as more knowledgeable (i.e., adults, a competent target, a supernatural being) makes the FB task harder to pass for children in the sensitive age-range. In this line, a paradigm that directly addressed children's inference of an informant's knowledge, found that 3-year-olds were more likely to attribute knowledge to smart characters that had not seen the content of a box versus non-smart characters that had seen the content (Lane et al. 2013). They found that this effect also extended to two other traits that should be less epistemically relevant (honesty and kindness), that is, most 3-year-olds used the valence of the informant's traits instead of the informant's perceptual access to the box's contents to infer their knowledge. The authors also assessed whether the participants trusted the testimony of the informant after having established whether the informant was knowledgeable or not. As would be expected, they found that when children incorrectly assumed that the nice and honest characters knew the content of the box, they also trusted their testimony. In the cases where the participant correctly assessed that the character did not know the content of the box, they did not trust the character's testimony. Interestingly, this pattern was not shared for the epistemic trait, the smart informants. Children that inferred that the smart informant did not know the content of the box still endorsed the smart's testimony versus the non-smart testimony. The researchers concluded that an informant's general intelligence might outweigh the informant's specific knowledge.

---

[4] To pass a FB task the participant needs to attribute a false belief to the target. Thus, children were more likely to pass the task with a peer target than with an adult target.

In the above studies the targets' identity influenced children's performance but it was not presented to the participants through a group membership lens, that is, the targets were not described as being an ingroup or an outgroup to the participant. In Seehagen et al. (2018) the peer target could be understood as an ingroup target (versus the adult [outgroup] target) and in Lane et al. (2010) the human target could also be understood as an ingroup target (versus the supernatural [outgroup] target) but neither manipulation specifically emphasized the distinction of the target's group membership in relation to the participant.

At the inception and design of the present research endeavor, there were no published developmental studies that specifically tested the influence of the protagonist's group membership in relation to the participant's group membership in FB paradigms. As of today, there are two studies that have addressed this question. In the first study, Sudo and Farrar (2020) tested the target's group membership influence in 3-to-5-year-olds' performance in the two standard FB tasks (among others). They conducted a between-group study with two manipulations of the target's group membership; a MGP manipulation and a cultural group manipulation in which they adapted the geographic manipulation by McLoughlin and Over (2017).[5] The target's ingroup or outgroup status was also manipulated as a between-subjects variable. The researchers found a significant interaction between children's age, degree of ingroup affiliation and target's group membership such that the older 4-year-olds and the 5-year-olds (58.62–70.40 months) who demonstrated higher ingroup affiliation were more likely to attribute false beliefs to an outgroup target than an ingroup target (i.e., performed worse in FB tasks with an ingroup target versus an outgroup target). Moreover, exploratory analysis showed older 4-year-olds and 5-year-olds performed worse at the tasks if they demonstrated higher degrees of affiliation with the target, regardless of the target's ingroup or outgroup status.

In the second study, Witt et al. (2022) conducted three experiments testing 4-year-olds' between-group performance in a change-of-location FB task using accent as the distinguishing social category between ingroup and outgroup targets. In the first experiment they found no difference in performance between the two groups. Exploratory analysis

---

[5] An ingroup character was described as someone who "lives around here, lives in a house just like yours, goes to a school just like yours, talks like you do, and wears normal clothes like you do." (Sudo & Farrar, 2020, p. 4). An outgroup character lived far away, in a different type of house, went to a different school, talked, and dressed differently.

revealed a significant effect of group membership (worse performance with an ingroup target) when they only included the participants that had correctly identified the protagonist's group membership. Experiments two and three aimed to replicate the significant effect but yielded no positive results. They conducted a fourth experiment in which they tested children in a battery of ToM tasks that included an unexpected-contents FB task. This time they used gender as the parsing social category. As in the other three studies, the authors reported no differences in performance between groups.

I would incur in an anachronic violation if I were to present the following hypothesis-driven research effort in the light of the results from the two studies presented above because they were not published when the author designed and conducted the experiments that will be presented in this work. Nonetheless, it will be crucial to compare their procedures and findings in the discussion section.

***

A few years ago, two separate articles planted the seeds that would inspire the current proposal. The first seed was a question from the *future research* section in Birch and Bloom's (2004) paper about the curse of knowledge: "how do our assumptions of how similar someone else is to us, for example in age, gender, or ethnicity, affect our assessments of whether that person will share our knowledge?" (Birch & Bloom, 2004, p.259). The second seed was an open call to research that was put forward in Rakoczy's (2014) brief commentary on Abrams et al.' 2014 article published at the British Journal of Developmental Psychology:

> "What these questions clearly point to is the need for a more systematic research programme on the relation of social cognition in the sense of thinking about individuals as members of certain social groups, on the one hand, and social cognition in the sense of ascribing subjective mental states to these individuals (ToM), on the other hand. The line of research to which the present paper belongs, investigating one such link between the two forms of social cognition (in which respects some forms of ToM might be prerequisites for complex group thinking) could be one of the many starting points for such a bigger project" (p.256).

The current proposal is, to my knowledge, the first pragmatic account to approach the cognitive branch of Developmental Intergroup ToM. On the one hand, it is a pragmatic undertaking because it stems from the premise that performance in FB tasks is not strictly linked to children's competence but rather intertwined with many pragmatic factors that make it cognitively demanding and context contingent. On the other hand, it belongs to the

cognitive branch of Developmental Intergroup ToM because it is concerned with the cognitive implications that result from the interaction between social intergroup biases (e.g., ingroup bias, mere membership bias) and task-related cognitive biases (e.g., curse of knowledge, egocentric bias) in FB task performance.

The pragmatic factor that will be manipulated in the current work is the target of the mentalizing effort in a FB task, also referred to as the protagonist. Success in these tasks is determined by the ability to register the difference between one's own experience and the protagonist's experience. The protagonist is the "false-belief holder" or the recipient of the participant's mentalizing projection and it differs from the participant in that they are ignorant to the situational knowledge the participant has. Because the understanding of the protagonist's "otherness" (distinct from "self") and therefore the protagonist's lack of access to the situational knowledge is indispensable to passing the task, the stepping point of my proposal is to enhance the "otherness" by presenting an outgroup protagonist. The current approach consists in manipulating the identity of the protagonist as a pragmatic factor that might enhance or worsen children's performance in their false belief attribution.

Drawing from the learnings in social cognition presented in the previous chapter, some assumptions can be made on how an outgroup protagonist might be perceived:

a) An outgroup protagonist will be categorized as an outgroup individual. Even if the task does not introduce the protagonist as an outgroup protagonist, children might categorize them as such. b) An outgroup protagonist will be considered dissimilar to oneself. c) An outgroup protagonist will be stereotyped as an outgroup individual. d) An outgroup protagonist will be valued less than an ingroup protagonist in warmth and in competence. e) An outgroup protagonist will be less preferred than an ingroup protagonist. f) An outgroup protagonist will not share one's knowledge. g) Children will be less likely to want to help an outgroup protagonist. h) All of the assumptions should also apply in artificially created groups.

Considering the pragmatic accounts and their explanations for FB task failure, the assumptions on how an outgroup protagonist might influence performance would be:

a) An outgroup protagonist will make the question more relevant and thus less pragmatically confusing. If the outgroup protagonist is not assumed to share one's knowledge, the question will deem as more plausible. b) Similarly, an outgroup protagonist will legitimize the experimenter's interest in the protagonist's beliefs. c) An outgroup protagonist will make the

question less academic. d) An outgroup protagonist will be less of a trigger for a helpfulness cooperative interpretation. e) An outgroup protagonist will help the participant to stay focused on the protagonist's perspective.

Lastly, based on the interpersonal cognitive biases that come into play during a FB task, some more assumptions (in order of appearance) on how an outgroup protagonist might influence performance can be done:

a) Children will be less biased by their own curse of knowledge when mentalizing about an outgroup protagonist. b) An outgroup protagonist will tilt the balance between one's perspective and the other's perspective causing sort of an altercentric bias. c) A mentalization about an outgroup protagonist will be less prone to egocentric distortion. d) Children will not overestimate the outgroup protagonist's knowledge. e) Children will not recruit the "self" template and project instead a "stereotype" ignorant template. f) Children will project low-competence and ignorance traits to an outgroup protagonist. Or, reversely, children will project high-competence and warmth traits to an ingroup protagonist with two immediate consequences: they might want their ingroup to succeed (strengthening the cooperative bias) and they might expect their ingroup to share their knowledge.

The combination and synthesis of all the assumptions above brings me to the following hypothesis to be proved or rebutted through the current experimental research: the outgroup protagonist functions as a cognitive heuristic (i.e., short-cut) for false-belief attribution. An outgroup protagonist reduces the processing cognitive load of the FB task because it has a direct effect on three pragmatic factors that influence performance: 1) it increases the saliency of the protagonist's perspective and 2) it creates an expectation of the protagonist's ignorance 3) it clarifies the experimenter's intent.

The salience of the protagonist's mental state is one of the variables that was found in Wellman et al. (2001)'s meta-analysis to have an influence on children's performance in FB tasks[6]. I predict that, by presenting a protagonist that is dissimilar to the participant, the saliency of the protagonist's perspective will automatically increase. The increased saliency will in turn help children stay focused on the perspective-tracking effort which is a key component of passing a FB task according to Rubio-Fernández and Geurts (2013).

---

[6] The other three being "presence" of the object in the test phase, "motive" (deceptive or not), and "participation" in the deception.

Moreover, under the light of the interpersonal cognitive biases, a target that is more salient will be less triggering of the egocentric biases that are making children fail the task (Birch & Bloom, 2004; Flavell, 1977; Nickerson, 1999; Robbins & Krueger, 2005).

The expectation of the protagonist's ignorance is another pragmatic factor that I predict will make the task easier for participants. Resorting to Ames' (2004) account on the double inferential mechanism for mind reading (projection from self-template versus stereotyping) and Westra's (2019) account on the relationship between stereotypes and mindreading, I expect that children that are presented with an outgroup protagonist will first project a competence-related stereotype onto the protagonist that will influence the participant's mindreading effort. Moreover, drawing from past experimental evidence with "knowledgeable targets" (i.e., adults – Seehagen et al., 2018; competent targets – Zmyj and Seehagen, 2020; a supernatural being – Lane et al., 2010) I expect a knowledgeable target to make the task more difficult than an ignorant one. If the outgroup protagonist is perceived as less competent than an ingroup protagonist, it will be easier for children to attribute a false belief to the outgroup protagonist. Further, the participant might also expect that an ingroup target would be more likely to share the participant's privileged knowledge than an outgroup target. That is, an intra-group relationship would assume knowledgeability whereas an inter-group relationship would assume ignorance. Perceiving the protagonist as less competent and less likely to share the participant's privileged ingroup knowledge should enhance children's performance in the FB task.

The third pragmatic factor is derived from the previous two factors. An outgroup protagonist whose mental states are more salient and who is perceived as less knowledgeable might reduce the pragmatic confusion of the test question thereby clarifying the experimenter's intent. Instead of interpreting the question in a normative way of what the response "should" be, an outgroup target might help children interpret the experimenter's intent as stemming from curiosity and not a teaching attempt.

It is important to note that the aim of the current account that will be referred to as the Outgroup Advantage Pragmatic Account (OAPA) is not to defend an early-belief ascription position nor a developmental conceptual shift approach. The goal of the account is to highlight some of the pragmatic performance issues in the FB task that muddle the distinction between children's competence and performance in false-belief attribution during the sensitive developmental years.

2.2. Research aims

- **Aim 1. To explore the influence of the protagonist's social group membership in false-belief performance in 4-to-5-year-olds.**
  Chapter 3. Pragmatic manipulation of the protagonist's group membership.

  o In Study 1, in a mixed-groups design, I will manipulate the protagonist's social membership resorting to a real social category (race). Half of the participants will be shown an ingroup target and the other half will be shown an outgroup target. I will first assess children's group membership preferences with explicit and implicit measures to make sure children are indeed categorizing the outgroup as such and show an ingroup bias. I will test participants in a five-item ToM scale to better frame the participant's performance in an unexpected-contents FB task by comparing it to other ToM tasks that develop during the preschool years.

  o In Study 2, in a mixed-groups design, I will define the protagonist's social membership via a MGP manipulation. Half of them will mentalize about a member of their team and the other half will do so for a member of the other team. The participants will complete the five-item ToM scale and an explicit measure for social preferences.

- **Aim 2. To explore the influence of the saliency of the protagonist's perspective in false-belief performance and how it interacts with group membership in 4-to-5-year-olds.**
  Chapter 4. Pragmatic manipulation of the saliency of the protagonist's perspective.

  o In Study 3, I will test children with a repeated-measures design with two versions of the unexpected-contents FB task: a neutral standard version and a version that will increase the saliency of the protagonist's perspective. In the second version the target will be presented as urgently needing the object that should be in the box instead of the unexpected content.

  o In Study 4, in a between-groups design, I will increase the saliency of the protagonist's perspective through the protagonist's membership manipulation of Study 2 and the object urgency manipulation in Study 3.

- **Aim 3. To explore the influence of the protagonist's expected ignorance in false-belief performance in 5-to-6-year-olds with a continuous measure.**
  Chapter 5. Pragmatic manipulation of the protagonist's expected ignorance.

  o In Study 5, in a mixed-design, I will test children on an adaptation of the change-of-location task that evaluates children's performance with a

continuous measurement paradigm. The protagonist's group membership and the group membership of the agent that misplaces the object will be determined by a MGP manipulation. Children will be assigned to one of four conditions: ingroup-ingroup, ingroup-outgroup, outgroup-ingroup, outgroup-outgroup.

- Aim 4. To explore the influence of the protagonist's expected ignorance in false-belief performance with an older population that is already proficient in standard FB tasks, with a continuous measure.

Chapter 6. Developmental trajectory of the pragmatic manipulation of the protagonist's expected ignorance

- o In Studies 6, in a mixed-design, I will test 8-to-9-year-olds following the same procedure as in Study 5.
- o In Study 7, in a mixed-design, I will test adults following the same procedure as in Studies 5 and 6.

Chapter 3.
Pragmatic manipulation of the protagonist's group membership

3.1. Study 1

3.1.1. Introduction

The pragmatic manipulation goal for this first study was to increase the likelihood of passing the FB task by underscoring the difference in participant-protagonist experiences by accentuating the "otherness" of the protagonist.

"Race,"one of the most salient social categories during infancy (Rhodes & Mandalaywala, 2017) was the chosen category to establish the "otherness" of the target. This choice allowed the study to rely on perceptual physical differences to trigger group categorization, without the need to add verbal descriptive group information that could increase the task's cognitive processing load. Taking into consideration the importance of the contextual experience of a child and their exposure to members of other races, African American children were selected as the outgroup protagonists of the mentalizing tasks, a social group with whom the Caucasian participants recruited would have had very little exposure.

Though the introductory chapters of the present research mainly focused on the FB task as the litmus test for ToM, it was also noted that ToM encompasses multiple other processes beyond false-belief attribution (Karmakar & Dogra, 2019). Specifically, Wellman and Liu (2004) reviewed all the tasks that had been used to assess ToM development up to that point and created a scale of five ToM items that tapped different facets of mental state understanding. The five-item ToM battery included the following tasks: 1) Diverse Desires (DD), 2) Diverse Beliefs (DB), 3) False Belief (FB), 4) Knowledge Access (KA), and 5) Real-Apparent Emotion (RAE)[7]. These tasks were ordered by difficulty in a developmental progression DD>DB>KA>FB>RAE. Since then, this scale has been used in numerous research endeavors across the world (e.g., Baker et al., 2021; De Gracia et al., 2020; Duh et al., 2016; Kristen et al., 2006; Pan Ding et al., 2021; Peterson & Wellman, 2009; Kuntoro et al., 2017; Selçuk et al., 2018; Shahaeian et al., 2011; Slaughter & Perez-Zapata, 2014; Toyama, 2007; Wellman et al., 2006).

---

[7] The Real Apparent Emotion task (RAE) is referred in some studies (Baker et al., 2021; Duh et al., 2016) as the Hidden Emotion task (HE).

In general terms, the first two tasks (DD, DB) put in conflict the participant's desires and beliefs with those of the protagonist. The two middle tasks (KA, FB) pit the participant's perspective and the protagonist's perspective against each other, with the former bearing real-world information that is inaccessible to the latter. Finally, the last task requires the hardest mentalization effort and it tests cognitive empathy skills.

Analyzing ToM skills that are framed within a sequence provides insight into the necessary steps to acquire each skill. If the order of the items is stable it means that the latter evolves from the former, that is, one skill lays the foundation of another skill (Duh et al., 2016). A sequential order that has a universal application would imply that there are "biological mechanisms of change" in play and a sequence that is conditioned by cultural factors would point to "environmental mechanisms" that affect performance (Duh et al., 2016, p. 583).

Interestingly, it can be observed that the sequence in which children pass these tasks is contingent to the cultural background of the participant, with children from more individualistic cultures performing in line with the original order of passing of the scale DD>DB>KA>FB>RAE (US: Peterson & Wellman, 2009 / Australia: Slaughter & Perez-Zapata, 2015 / Germany: Kristen et al., 2006 / Poland: Smogorzewska et al., 2018) and children from more collectivist cultures passing the Knowledge Access task before the Diverse Beliefs task (Iran: Shahaeian et al., 2011 / China: Wellman et al., 2006, Duh et al., 2016 / Turkey: Selçuk et al., 2018 / Japan: Toyama, 2007 / Singapore: Kuntoro et al., 2017).

The Knowledge Access task is the only task in the scale that requires a yes/no answer. Westra and Carruthers (2017) explain that the cross-cultural difference in performance can be driven by the experimental artifact alone that triggers children's yes-bias[8]. Since the task's success depends on giving two negative answers, being able to inhibit the yes-bias is crucial for passing the task and is strongly predicted by the child's command in executive functions, specifically inhibitory-control. Hence, the cross-cultural performance difference relies on cross-cultural cognitive developmental differences in inhibitory control, with children from more collectivist cultures mastering inhibitory control skills earlier than their individualistic-cultures counterparts (Westra & Carruthers, 2017). A complementary view is based on

---

[8] The authors refer to the work of Fritzley and Lee (2003) and Okanda and Itakura (2008) that shows children's tendency to answer positively to all questions.

cultural factors like conversational styles and socialization processes and states that collectivist cultures put more emphasis on the weight of knowledge and respecting those that have pragmatic knowledge (Duh et al., 2016; Shahaeian et al., 2011; Wellman et al., 2006) which can lead to children's responding to the question from a "seeing leads to knowing" point of view, instead of focusing on the item in the box (Westra & Carruthers, 2017). And yet another factor to consider is that children in individualist cultures are more proficient at the DB task and learn to pass it before the KA task because their cultures reward the expression of different opinions and beliefs (Westra & Carruthers, 2017).

The widespread use of this paradigm and the nuances in pragmatic understanding contingent upon cultural differences seemed like a very interesting approach to frame children's FB task performance among other tasks that could potentially be also pragmatically affected by the identity of the protagonist.

To complement the ToM research effort and provide in-study evidence of children's social preferences, one measure of implicit social preferences (the Implicit Racial Bias Test "IRBT", Qian & Lee, 2016) and one measure of explicit social preferences (the Explicit Preferences Task, Qian & Lee, 2016) were included.

Two control measures were also included: the Spanish adaptation of the Peabody Picture Vocabulary Test (PPVT) (Dunn, Dunn & Arribas, 2006) as a between-group control for semantic knowledge and the Racial Categorization Task (Aboud & Mitchell, 1977) to discard children that were not able to racially categorize the depicted protagonists.

In a between-group design, I tested an all-Caucasian 4-to-5-year-olds' sample with African American protagonists (outgroup) and Caucasian protagonists (ingroup).

Attending to the primary goal of this study, the first prediction was that performance in the FB task would be significantly higher in the racial outgroup condition. In an unexpected-contents FB task children need to compute three different perspectives: 1. Prior self-perspective that is based on common knowledge (there are band-aids in a band-aid box) 2. Updated self-perspective shared with the experimenter (there are crayons in a band-aid box) 3. The protagonist's perspective that is ignorant to the situational knowledge (there are band-aids in a band-aid box). Though the participant could project their own prior ignorance to

the protagonist, the weight of knowing the actual content of the box added to the fact that this knowledge is also shared by the experimenter makes children struggle to respond from the protagonist's point of view.

As outlined in the introduction, the OAPA points to three pragmatic factors that might make a task less cognitively demanding with an outgroup protagonist: the saliency of the protagonist's perspective, the protagonist's expected ignorance and the experimenter's intent.

In the unexpected-contents task, the experimenter's communicative action (their question) might generate a normative bias (Helming et al., 2016). The set-up of the task, asking the child about the content of the box, can be conceived as a trick. The child answers "smarties" and feels tricked when they see there are "crayons" in the box. The reveal of the true content of the box can be perceived as a teaching demonstration and the experimenter's new question post reveal can then be understood as a warning signal to not be fooled again. The motivation to show their knowledge and avoid being tricked again might make them interpret the prediction question into a normative question. I predicted that an outgroup protagonist whose perspective would be more salient and who would also be perceived as less knowledgeable (less competent and with no access to the participant's privileged knowledge) would help children understand the test question as a legitimately curious question instead of a teaching demonstration.

Since it was decided to frame the FB task's performance within a five-item ToM scale, I also attempted to predict children's performance for the remaining four tasks. The predictions were derived from Westra and Carruthers' (2017) pragmatic review on Wellman and Liu's mindreading scale.

Regarding the Diverse Desires task's performance, although the OAPA would predict that children would be more likely to be biased by their own preferences with an ingroup protagonist, Westra and Carruthers (2017) state that children have wide experience with conversations about desires from a very early age and argue that there is really "no age at which infants are truly egocentric about desires" (p.172). Adding this knowledge to the fact that it is the easiest task in the scale and the scores are usually high, a weak outgroup advantage (if any) was predicted.

As for the Diverse Beliefs task, it was predicted that children's performance would also not significantly improve with an outgroup protagonist. From a pragmatic perspective this task is easier than the FB task because the protagonist' view is expressed by the experimenter after the child has shared their view and it can be interpreted as if the experimenter is providing the correct answer and encouraging the participant to do so (Westra & Carruthers, 2017). If the experimenter's view overrides the participant's view, there is again slight room for improvement.

Regarding the Knowledge Access task, two predictions were made for this task. First, in line with the FB task prediction, children in the outgroup condition would perform better than their ingroup counterparts. By projecting their knowledge onto the similar protagonist, it was expected that participants in the ingroup condition would make more mistakes. The second KA prediction referred to the tasks' position in the scale's difficulty progression.

To determine whether the country of testing (Spain) was closer to the more individualistic cultures or the collectivist ones, I resorted to Hofstede's 100-point scale of individualism and collectivism (Hofstede Insights, 2018) following the procedure in Huppert et al.'s (2018) extensive cross-cultural study. The scale showed that Spain falls in the middle of the individualist-collectivist dimension with a 51% score in individualism. This score was then compared to the countries that had been previously grouped in the individualistic side (US: 91%, Australia: 90%, Germany: 67%) and the countries that had been grouped in the collectivist side (Iran: 41%; Turkey: 37%, China: 20%). It seemed that Spanish children's performance should then slightly lean towards the sequence found in the collectivist cultures: DD>KA>DB>FB>RAE. I could predict the ingroup's sequential order based on performance in other countries because the ingroup set-up was framed like the standardized ToM scales that were performed in all these countries (without any pragmatic manipulation). Though there was no existing reference based on experimental data for the outgroup condition, it was predicted that the cultural-contingent effect (that made children succeed at KA before DB in collectivist cultures) should also apply with an outgroup protagonist.

As regards to the last task in the scale (RAE), it was predicted that children in the ingroup condition would be more likely to decipher the protagonist's emotions as the performance in this task should benefit from projecting one's experience onto the protagonist.

The prediction for the social preferences tasks was that all children would show implicit and explicit social preferences for the ingroup, and these preferences would not correlate with performance in the ToM five-item scale.

3.1.2. Method

Participants

The final sample consisted in 130 4-to-5-year-olds (67 females, M_age= 52.1 months, SD_age: 8 months) from two private schools in Barcelona with equivalent socio-economic status. I initially collected data from 84 4-to-5-year-old children (42 females, M_age= 53 months; SD_age: 7 months) from a private school in Barcelona. The analysis showed that there was a significant difference in FB performance between conditions. To ensure that the possible results were not due to limited power, I pre-registered a replication study and planned to increase the sample by 100 kids (50 per condition) which, on its own, would have had 80% power to detect a moderate effect size (d=0.5)[9]. Due to Covid-19 I was only able to test 46 participants (23 per condition). A 2 (condition) by 2 (school) ANOVA revealed a main effect of condition, $F(1,126) = 5.41$, $p = .022$, $\eta_p^2 = .04$, but no school effect, $F(1,126) = 0.68$, $p = .41$, $\eta_p^2 = .00$, or interaction effect , $F(1,126) = 0.55$, $p = .46$, $\eta_p^2 = .00$. The data from the two schools was merged into the final sample of 130 participants with a power of 88% to detect a medium size effect (d=0.5). The study was approved by the institutional review board of the university and a caregiver's informed consent was obtained for each participant.

[9] This power analysis was calculated with the software package G*Power (Faul et al., 2009)

Measures

Peabody Picture Vocabulary Test

The PPVT (Dunn, 1997) is a norm-referenced instrument that assesses receptive vocabulary of children and adults. The Spanish version of the test (Dunn et al. 2006) was administered with the stimulus booklet and was later coded using the standardized guidelines.

Coding. To assess semantic knowledge, the intellectual coefficient measure that is extracted by task performance and controls for participant's age was used.

Theory of Mind scale

Wellman and Liu's (2004) five-item Theory of Mind scale for preschoolers is composed of five different tasks that are set to vary in difficulty but are comparable in format (props and materials). In order of difficulty, the scale first requires children to theorize about desires and beliefs that are in conflict with one's own (Diverse Desires, Diverse Beliefs), then makes children assess a character's access to situational knowledge that the participant has (Knowledge Access) as well as a character's false belief in an unexpected-contents standard task (False Belief) and finally it requires children to theorize about a character's emotional response to an event (Real-Apparent Emotion). See Table 1 for the tasks' full description as reported in Wellman & Liu (2004) – these were also the exact versions that were used in the current study.

Table 1. Full description of each task in the five-item ToM scale (Wellman & Liu, 2004).

DD	Children see a picture of a child next to a picture of a carrot and a cookie. "Imagine that it's snack time and this kid wants a snack to eat. Here are two different snacks: a [carrot] and a [cookie]. Which snack would you like best? Would you like a [carrot] or a [cookie] best?" This is the own desire question. If the child chooses the carrot: "Well, that's a good choice, but this kid really likes [cookies]. He doesn't like [carrots]. What he likes best are [cookies]." (Or, if the child chooses the cookie, they are told the kid likes carrots.) Then the child is asked the test question: "So, now it's time to eat. This kid can only choose one snack, just one. Which snack will this kid choose? A [carrot] or a [cookie]?"
DB	Children see picture of a girl and a picture of a garage and a picture of bushes. "Imagine that this kid is looking for her cat. Her cat might be hiding in the [bushes] or it might be hiding in the [garage]. Where do you think the cat is? In the [bushes] or in the [garage]?" This is the own belief question. If the child chooses the bushes:

	"Well, that's a good idea, but she thinks her cat is in the [garage]." (Or, if the child chooses the garage, they are told the girl thinks her cat is in the bushes.) Then the child is asked the test question: "So where will she look for her cat? In the [bushes] or in the [garage]?".
KA	Children see a physical nondescript plastic box that contains a plastic dog toy figurine. "Here's a box. What do you think is inside the box?" (The child can give any answer they like or indicate that they do not know). Next, the box is opened, and the child is shown the content of the box: "Let's see, it's really a [dog] inside!" The box is closed: "Okay, what is in the box?" Then they see a picture of a girl on the screen: "This kid has never ever seen inside this drawer. Now here she comes. So, does she know what is in the box?
FB	The child sees a clearly identifiable Band-Aid box. "Here's a [Band-Aid box]. What do you think is inside the [Band-Aid box]?" Next, the Band-Aid box is opened: "Let's see, it's really [crayons] inside!" The Band-Aid box is closed: "Okay, what is in the [Band-Aid box]?" Then they see a picture of a boy on the screen: "This kid has never ever seen inside this [Band-Aid box]. Now here he comes. So, what does he think is in the box? [Band-Aids] or [crayons]?
RAE	Initially, children see a slide with three cartoon-like drawings of faces on it. A happy face, a neutral face, and a sad face. Then the child is shown a picture of a kid standing backward so their facial expression cannot be seen. "This story is about a kid. I'm going to ask you about how the kid really feels inside and how he looks on his face. He might really feel one way inside but look a different way on his face. Or he might really feel the same way inside as he looks on his face. I want you to tell me how he really feels inside and how he looks on his face." "This kid's friends were playing together and telling jokes. One of the older children, told a mean joke about him and everyone laughed. Everyone thought it was very funny, but not him. But he didn't want the other children to see how he felt about the joke, because they would call him a baby. So, he tried to hide how he felt." Then the child gets two memory checks: "What did the other children do when the older kid told a mean joke about him?" (Answer: Laughed or thought it was funny.) "In the story, what would the other children do if they knew how he felt?" (Answer: Call him a baby or tease him.) Bring back the slide with the three emotion pictures: "So, how did he really feel, when everyone laughed? Did he feel happy, sad, or okay?" (The test *feel* question) "How did he try to look on his face, when everyone laughed? Did he look happy, sad, or okay? (The test *look* question).

Instead of using cartoon characters, each task was paired with a racial ingroup (Caucasian) or outgroup (African American) protagonist that was gender-matched to the participant. See examples of the DD task in Figure 3 and the DB task in Figure 4. Sixteen facial pictures of real children with "happy" expressions (four by racial group per gender) were selected from the Child Affective Facial Expression set (CAFE, LoBue, 2014) and were counter-balanced across the tasks. Protagonists were referred to in a neutral manner (e.g., "this kid") without

any cultural or geographical identification. Two versions of each task were made to be counterbalanced across participants and control for potential unwarranted saliency of the chosen object. Location of the correct object was also counterbalanced. The tasks were administered in order of difficulty as seen in previous literature.
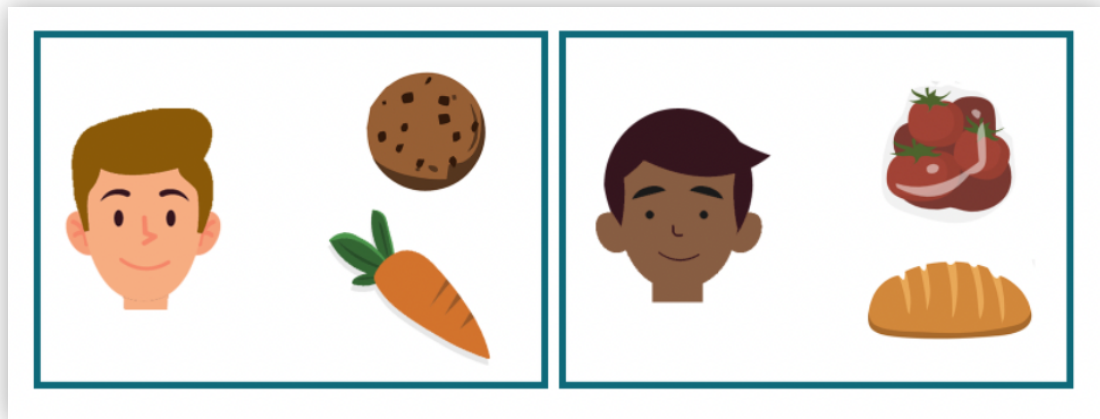
Figure 3. Recreation of DD with an ingroup target with version A's food items (left) and an outgroup target with version B food items (right). The original children's images from the CAFE dataset are copyright protected, this image is a recreation of the stimuli that was specifically illustrated for this dissertation.
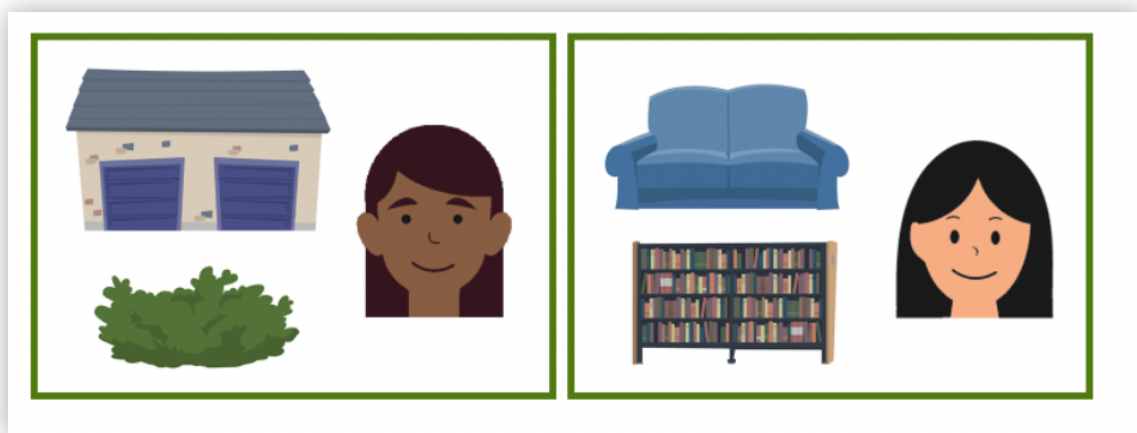
Figure 4. Recreation of DB with an outgroup target with version A's locations (left) and an ingroup target with version B's locations (right).

Coding. Wellman and Liu (2004) established a 5-point dichotomic scoring system to test children's performance in their five-item ToM scale. Children are assigned 1 point if they pass and 0 if they fail, resulting in a max. score of 5 and a min. score of 0. In Wellman and Liu's (2004) scoring system, in the KA and FB tasks, children were only assigned 1 point if they were able to correctly answer a memory control question that was asked right *after* the test question ("Did she see inside the box?"). Pilot testing showed that most children would

give the same answer to test and memory questions in the KA task (incorrect: yes/yes; correct: no/no) and would consistently say "yes" (incorrect) in the memory FB question, even if they had given the correct response to the test question. Similar effects were seen in Rubio-Fernández and Geurts (2015) where children gave the same response to the check-in question "What's in here?" and the test question of the unexpected-contents FB task. Rubio-Fernández (2017) suggests that this pattern of responses can be related to the type of perseverative errors detected by Deák and Enright (2006) on the Appearance/Reality task. According to the Pragmatic of Questions Account (Déak et al., 2003), preschoolers might not appreciate that when asked multiple questions about the same topic, each question requires a different answer. It is the lack of pragmatic understanding that "leaks" (p.951) responses across questions. In the actual test the memory control question was removed from the KA and FB trials.

To my knowledge, there had not been any previous research that investigated performance with an outgroup protagonist using the full five-item scale, thus I was interested in analyzing performance by the overall 5-point scale as well as by individual tasks, specifically focusing on the FB task. To see whether the group manipulation influenced the tasks' sequence of difficulty, the Guttman and Rasch scale analyses, which were the scaling models that Wellman and Liu (2004) used to create the scale based on individual performance patterns, were applied.

Implicit Racial Bias Test

The Implicit Racial Bias Test (IRBT) was developed by Qian et al. (2016) and it offers a window into very young children's implicit preferences for their racial ingroup. It is a child-friendly adaptation of the widely used Implicit Association Test (IAT, Greenwald et al., 1998) that assesses the speed with which a participant associates positive and negative attributes (concepts) to a racial ingroup character or a racial outgroup character. The IRBT also measures response latencies of a quick target association, but it replaces the concepts with cartoons of a smiley face and a frowny face to reduce cognitive demand.

I was able to use the original Black - White IRBT stimuli (Qian et al., 2016) that was kindly provided by the authors. See an example in Figure 5. The test includes 20 Black faces (10 per

gender) and 20 Caucasian faces (that had previously been rated for attractiveness and distinctiveness).
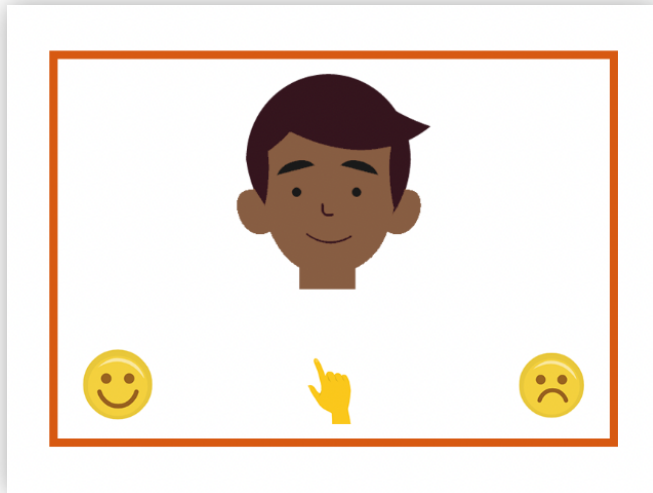
Figure 5. Recreation of the participant's view of an IRBT outgroup trial. Following the instructions, in the congruent block, the child would need to press the frowny face. In the incongruent block the child would need to press the smiley face.

On each trial the participant saw a face in the middle of the screen (17cmx21cm) with a cartoon of a smiley and a frowny face placed near the bottom at the two sides of the screen. There were 8 practice trials, and 40 test trials divided into two blocks: congruent trials and incongruent trials. In congruent trials participants were instructed to press the smiley icon when they saw an ingroup face and the frowny icon when they saw an outgroup face. In incongruent trials they were instructed to press the smiley icon for the outgroup faces and the frowny icon for the ingroup faces. The sequence of congruent vs. incongruent blocks and the icon position was counterbalanced across participants.

Coding. To measure implicit bias, the authors' guidelines were followed, and a *D score* was obtained for each child by calculating the difference between the average response latencies between racial conditions and dividing it by the standard deviation of response latencies across conditions.

Explicit Preferences Task

An adaptation of the Explicit Preferences Task (Qian et al., 2016; Setoh et al., 2019) was administered. In this task, children had to choose between a racial ingroup member and a

racial outgroup member in five different scenarios (e.g., Imagine that you are going to the pool with your family. Who would you like to come with you? This kid or this kid?). The images were taken from the CAFE Child Affective Facial Expression set (CAFE, LoBue, 2014). Stimuli was gender-matched to the participant and the position of the characters was counterbalanced across trials. See a male and female trial examples in Figure 6.
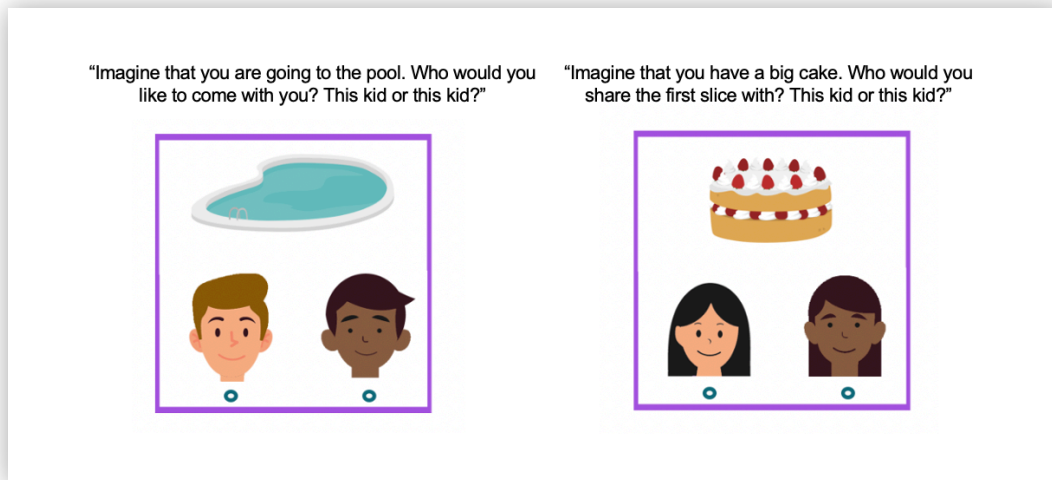
Figure 6. Recreation of a male trial (left) and a female trial (right) of the Explicit Social Preferences Task.

Coding. Trials were scored as with one point if they chose an ingroup member and zero if they chose an outgroup member. Scores were calculated as the proportions of trials where the participant chooses the ingroup over the outgroup.

The Racial Categorization Task

The Racial Categorization Task (Aboud & Mitchell, 1977) was merely used as an end-of-study control task to make sure that children were able to differentiate between the racially distinct protagonists. Eight pictures of child faces from the CAFE Child Affective Facial Expression set (CAFE, LoBue, 2014) were presented on a slide and participant were asked to point to Black individuals or White individuals. All children were able to point at the correct target in four consecutive trials and no participants had to be discarded.

Procedure

All participants were tested in-person during 30 min individual sessions with an experimenter and a research assistant in a quiet room of the school. A Macbook Pro laptop connected to a 16" ViewSonic TD1630-3 touch screen monitor was used for all the computerized tests. The ToM Scale (Wellman & Liu, 2004), the Explicit Preferences Task (Qian et al., 2016; Setoh et al., 2019) and the Social Categorization task (Aboud & Mitchell, 1977) were presented in Qualtrics software (Qualtrics, Provo, UT). The IRBT (Qian et al., 2016) response latencies were collected with E-prime 2.0 Professional (Psychology Software Tools, Sharpsburg, PA). The PPVT (Dunn, Dunn & Arribas, 2006) was presented in its printed version with the official stimulus booklet. The order of the tasks remained the same across participants, the PPVT was presented first to help participants get familiarized with the experimenters and was followed by the ToM Scale, the IRBT, the Explicit Preferences Task and the Social Categorization task. Participants were tested in their native language, Catalan[10].

3.1.3. Results

ToM scale

The between-group performance was first analyzed by comparing the means of the complete Wellman and Liu's five-item ToM scale (0-5 points) as seen in previous literature (e.g., Wellman et al., 2006). I ran a moderated regression analysis with the complete five-item scale as the dependent variable, two predictors (group condition and ingroup explicit preference) and one control variable (school). The linear regression model showed that children's performance was predicted by group condition, $t(125)=2.61$, p < .01, but not by ingroup explicit preference, $t(125)=0.40$, $p = 0.67$, or by school, $t(125)=0.25$, $p = 0.78$.

A Welch two-sample t-test confirmed that children in the outgroup condition (M=3.2) outperformed children in the ingroup condition (M=2.7), $t(121) = -2.33$, $p = .02$, d=0.41. Further, analyzing each task individually showed that Knowledge Access and unexpected-contents False Belief were the two tasks that mainly drove the overall between-group difference. In both tasks the outgroup condition significantly outperformed the ingroup condition (Knowledge Access: $t(114) = -2.80$, $p = .006$, d=0.49/ Unexpected-contents False

---

[10] See the Catalan translations of the five tasks in the Wellman and Liu's (2004) ToM scale in the Appendix.

Belief: $t(126) = -2.73$, $p = .007$, $d=0.48$). While in the Knowledge Access task both groups' performances were significantly above-chance, in the FB task only the outgroup condition performed above-chance, $t(64) = 3.65$, $p < .001$, $d=0.45$, compared to the ingroup's at-chance performance, $t(64) = 0.37$, $p = .713$, $d=0.05$.

Table 2. Percentages of participants passing each task in the ToM scale and TOM overall score by condition.

ToM Task	Ingroup condition $N =65$	Outgroup condition $N =65$
Diverse Desires	72%	67%
Diverse Beliefs	61%	72%
Knowledge Access	67%	87%
False Belief	48%	70%
Real-Apparent Emotion	23%	26%
ToM five-item overall score	**2.72**	**3.25**

Order by difficulty in the ToM scale

Table 2 shows the percentages of participants passing each task. Descriptively, children in the ingroup condition followed the pattern of difficulty that had been found in collectivist cultures (Baker et al., 2021) where Knowledge Access moves up one step to pass Diverse Beliefs: DD > KA > DB > FB > RAE. Children in the outgroup condition seemed to follow a completely new pattern.

Following previous analyses of Wellman & Liu's five-item ToM scale (Baker et al. 2021; Duh et al., 2016; Wellman & Liu, 2004; Peterson & Wellman, 2014), each condition's sequence of ToM tasks were examined using two models of scale analysis on individual performance patterns: the Rasch Model (Rasch, 1960) and the Guttman Model (Guttman, 1944, 1950). Both models aim to describe children's response patterns based on respondent's ability and task difficulty, but they do so through diverging item-response functions. While for a Guttman scale the item-response functions are deterministic, the item-response functions for a Rasch model are probabilistic (Wellman & Liu, 2004).

As detailed in Duh et al. (2016), the Guttman model organizes the tasks by difficulty in a strict sequential order which means that a well-fitted scalogram will only take place when the items significantly vary in difficulty. With five tasks, the model fit increases when a participant's response exhibits one of the six patterns in Table 3.

Table 3. Guttman Scalogram patterns for the five-item ToM scale by condition.

ToM task	Pattern						Other patterns	N
	1	2	3	4	5	6		
Diverse Desires	-	+	+	+	+	+		
Diverse Beliefs	-	-	+	+	+	+		
Knowledge Access	-	-	-	+	+	+		
False Belief	-	-	-	-	+	+		
R-A Emotion	-	-	-	-	-	+		
Ingroup condition	3	7	5	7	13	7	23	65
Outgroup condition	1	1	2	4	20	6	31	65
Total	4	8	7	11	33	13	54	130

The Gutman model analysis revealed that the developmental sequence that had been found in individualist cultures (DD > DB > KA > FB > RAE) captured 65% of the ingroup responses with a coefficient of reproducibility of .82[11] and 52% of the outgroup responses with a coefficient of .76. On the other hand, the scale found in more collectivist cultures (DD > KA > DB > FB > RAE) captured 58% of the ingroup responses with a coefficient of reproducibility of .82 and 51% of the outgroup responses with a coefficient of .80.

Guttman's scale inability to capture items of similar difficulty can be less problematic in the Rasch Model (Wellman & Liu, 2004). In a dichotomous scored scale, the Rasch analysis offers less strict scale progressions as it models the probability; when a participant passes item N correctly (DB for example), it means that this participant probably also passes item N-1(DD) (Wellman & Liu, 2004). The Rasch model analysis for the five items was analyzed using the eRm package (Mair & Hatzinger, 2007) through R (R Core Team, 2022).

Tables 4 and 5 show the task difficulty levels as well as the model's item fit statistics for each condition. As detailed in Duh et al. (2016), there are two types of fit statistics that evaluate item fit of a Rasch model sequence to the actual data; infit and outfit, which are sensitive to unusual responses near the item and far from the item respectively (Wright & Masters, 1982). For standardized infit and outfit statistics, t-values that are over 2 indicate higher variability

---

[11] This coefficient was calculated using Green's (1956) estimation procedure and only .90 and above would be significant.

than expected and negative values over -2 indicate too much predictability (Wellman & Liu, 2004).

Table 4. Rasch Model item difficulty levels (easiest to hardest) and fit statistics for the ingroup condition.

ToM Task	Difficulty(eta)	SE	Outfit *t*	Infit *t*
Diverse Desires	-1.02	0.28	0.34	0.75
Knowledge Access	-0.75	0.27	-1.93	-2.06
Diverse Beliefs	-0.41	0.27	0.41	0.46
False Belief	0.36	0.27	-2.41	-2.29
R-A Emotion	1.82	0.33	0.00	0.28

Table 5. Rasch Model item difficulty levels (easiest to hardest) and fit statistics for the outgroup condition.

ToM Task	Difficulty(eta)	SE	Outfit *t*	Infit *t*
Knowledge Access	-1.39	0.35	-1.70	-1.59
Diverse Beliefs	-0.26	0.27	-1.82	-1.46
False Belief	-0.18	0.27	-1.10	-1.12
Diverse Desires	-0.02	0.27	-0.31	-0.38
R-A Emotion	1.85	0.28	2.34	0.79

In the ingroup condition (Table 4), DD, DB and RAE showed a perfect fit and KA and FB showed a slightly poorer fit[12]. In the outgroup condition (Table 5), all items except RAE (infit $t = 0.79$, outfit $t = 2.34$) showed a perfect fit.

Following Baker et al.'s (2021) Rasch model analysis the Item Characteristics Curves (ICC) was plotted, an S-shaped curve that shows the probability of passing an item as a function of ability for each task on the scale. The difficulty of the item is set at the point at which a participant has a 50% chance of passing the task successfully (Baker et al. 2021). On the X axis, the items are presented from easiest to hardest (Figure 7).

In the ingroup condition, the ICC shows the developmental sequence found in collectivist cultures where KA results easier than DB. The ICC for the outgroup condition shows a completely new pattern in which Knowledge Access, Diverse Beliefs and unexpected-contents False Belief appear easier than Diverse Desires.

---

[12] According to Linacre (2003), $t = 2 – 2.9$ is considered noticeable unpredictable and $t \geq 3$ would be unpredictable.
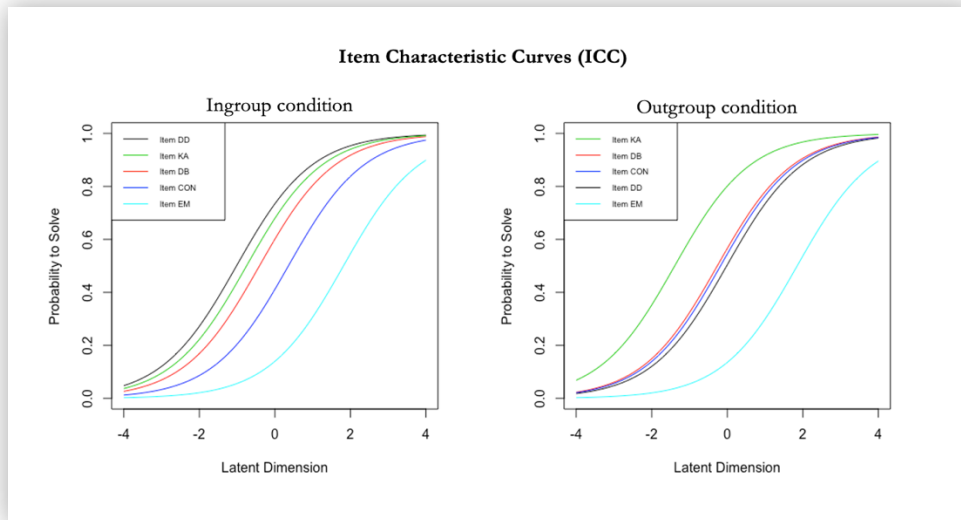
Figure 7. ICC for the ingroup condition (left) and the outgroup condition (right).

Rasch modeling shows that, on top of the overall difference in performance between the ingroup and the outgroup conditions, there was also a between-group difference in task difficulty. While the ingroup condition was closer to the sequence found in collectivist cultures, the outgroup condition showed an entirely new sequence.

ToM and explicit preferences

Data for the Explicit Preferences Task was obtained for the complete sample (N=130).

All children showed an explicit bias favoring their racial group, t (129)= 12.86,

$p < .001$, d=1.13, and the bias did not differ between conditions ($M\_ingroup = 0.76$, $M\_outgroup = 0.80$), $t(128)= 0.98$, $p = .33$, d=0.17. Moreover, a Pearson correlation analysis showed that explicit bias was not correlated to performance in the ToM scale, $t(128) = 0.75$, $p = .46$, $r =.07$.

ToM and IRBT

Data for the IRBT was obtained for 79 children from the first sample (5 participants failed to complete the task). Due to the long duration of this test, I refrained from using it for the second pool of participants.

All children showed an implicit bias favoring their racial group, t (78)= 4.61, $p < .001$, d=0.52, and the bias did not differ between conditions ($M\_ingroup = 0.19$, $M\_outgroup = 0.32$),

$t(76.87)= 1.16, p = .25, d=0.26$. Moreover, a Pearson correlation analysis showed that implicit bias was not correlated to performance in the ToM scale, $t(77) = 0.15$, $p = .88$, $r =0.02$.

ToM and PVVT

Data for the PPVT was obtained for the first 84 children in the sample. Due to the long duration of this test, I also refrained from using it for the second pool of participants.

No between-group differences were found in semantic knowledge ($M\_ingroup$ = 108.32, $M\_outgroup$ = 110.44), $t(76.46)= 0.72, p = .47, d=0.16$. This control measure allowed me to establish that the difference in the ToM scores between conditions was created by the pragmatic manipulation and not due to differing competence between the two groups.

3.1.4. Discussion

The main goal of this first study was to determine whether increasing the protagonist's otherness would influence children's performance in a FB task. To accomplish this, I conducted a between-groups pragmatic manipulation of the protagonist with the social category "race." Half of the participants mentalized about a racial ingroup protagonist and the other half mentalized about a racial outgroup protagonist. To assure that the chosen social category would be salient for children this age and trigger their social categorization processes, children's social preferences were tested with explicit and implicit measures. Moreover, in the interest of framing the FB task performance within a wider range of ToM tasks, 4-5-year-olds were tested with the five-item ToM scale by Wellman and Liu (2004) and predicted that the manipulation would influence the tasks on different degrees. Lastly, between-group differences in semantic knowledge were controlled for and an end of study inclusion question on racial categorization was added.

In line with the predictions of the OAPA, children's performance in the unexpected-contents FB task was significantly better in the outgroup condition. In this task, the child had to inhibit their own situational knowledge that was shared by the experimenter and mentalize from the protagonist's perspective assuming an ignorant point of view. My hypothesis was that an outgroup protagonist would make the task less cognitively demanding because it would reduce the pragmatic difficulty of the task. The three factors that I pointed to were the saliency of the protagonist's perspective, the protagonist's expected ignorance, and the experimenter's intent. Potentially, a combination of these factors was responsible for the

statistically significant superior performance of the outgroup condition in the FB task as well as the KA task. I anticipated performance in the KA task to mirror FB performance because the pragmatic factors that were manipulated applied equally to both. The high scoring of these two tasks lead to a significant superiority in the overall performance of the five-item scale score. As predicted, performance in DD and DB was not significantly different.

Contrary to the predictions, the present study showed no differences in performance in the RAE task. The RAE task is the only task in the scale that tests cognitive empathic understanding, and, according to the OAPA, it would be the only task in the scale that benefits from having an ingroup protagonist. Success in this task differs from the FB task because it does not pit the participant's perspective against the protagonist's, instead it requires that the participant can project how they would feel in the protagonist's situation. As reviewed in the general introduction, previous evidence in developmental research has shown that, in tasks where there are no conflicting perspectives, an ingroup protagonist promotes mental state attribution (e.g., Gönültaş et al., 2019; McLoughlin & Over, 2017). Children's scores for the RAE task were very low in both groups so there is the possibility that a difference between conditions could go unobserved due to a floor effect.

Crucial to the present research, it was found that the pragmatic manipulation brought children in the outgroup condition to perform above chance levels in the FB task (compared to the at-chance performance of the ingroup condition). This is important because, as reviewed in the introduction, Wellman et al.'s (2001) meta-analysis reported that children's performance could be improved with some tasks manipulations, but in no case would these manipulations lift performance to above-chance levels. The present finding adds experimental evidence to the pragmatic accounts that aim to explain how children's competence in false-belief attribution can be highjacked by pragmatic factors resulting in poor performance.

As regards to the developmental sequence of the five-item ToM scale, as had been anticipated, the pattern of responses in the ingroup condition was closer to the collectivist pattern (DD>KA>DB>FB>RAE) where KA moves up one step. Remarkably, the outgroup data did not fall into either the collectivist or the individualistic sequential pattern. Like in collectivist cultures and in this Study's ingroup condition, children were more likely to pass the KA before the DB. Nonetheless, the pragmatic manipulation made KA become even easier, making it the easiest task in the scale. Moreover, children's performance boost

in the FB task with an outgroup protagonist further upended the sequential order of the scale as children's FB passing scores became equivalent to their DD and DB scores.

The instability that was caused by the protagonist manipulation adds new experimental evidence to the debate on ToM skills acquisition beyond false-belief attribution. A robust sequence would point to a developmental scalable attainment and refinement of ToM concepts, one building off another. Though the universality of the scale has long been disproven from a cross-cultural perspective, the fact that it is only one of the tasks that switches places still allows for a constructivist understanding of ToM development.

However, the disruption of the sequential order that derived from manipulating the same pragmatic factor for all tasks in the present study, points to a more flexible understanding of ToM development that can be dependent on pragmatic factors that can mask children's competence.

Importantly, I was able to determine that the between-group difference in the scale's performance was due to the task manipulation and discard any between-group differences in cognitive abilities related to social categorization or language. On the social front, as expected, it was found that all participants showed an explicit and implicit ingroup bias and these were not correlated with performance. On the linguistic front an equivalent performance in semantic knowledge for both conditions was found.

In the subsequent study the aim was to gather further evidence to assess false-belief attribution performance with ingroup/outgroup protagonists created through a MGP assignment.

3.2. Study 2

3.2.1. Introduction

The purpose of this study was to replicate the pragmatic manipulation of Study 1 with an artificial social category created on the spot. Instead of using the social category "race," the protagonist's otherness was defined through a MGP assignment.

The reason an experimental paradigm was used to trigger the participant's social categorization processes is that it allows for a noiseless understanding of intergroup preferences, stripped from confounding factors such as exposure or pre-existing group-related stereotypes (Richter et al., 2016). The mere fact of belonging to a group has been widely shown to be enough to generate ingroup biases in multiple dimensions (Dunham, 2018a). Here, the aim was to see whether the MGP's powerful parsing quality would also apply in false-belief attribution.

To create the MGP, the procedure used in Richter et al. (2016) that is specifically targeted for young children was adapted. In this procedure children are introduced to two groups (identified by distinctive colors) and are asked to draw a colored coin from a bucket to see which group they belong to. Once they have been assigned a group, children are given group markers in the shape of armbands, stickers and scarves. In Richter's manipulation the two groups were described in a neutral manner and the participants were not given any descriptive information about them that would make them different. In this study, I decided to take advantage of using an artificial category and increased the ingroup/outgroup distinction by giving one item of information about character traits. Nesdale et al. (2005) found that ingroup identification in a MGP setting that only lasts the duration of the study is strengthened if the outgroup is perceived to be threatening. In this manipulation, I wanted to make ingroup identification powerful and effective without giving children excessive social information to process. Once they had been put in a group, they were told they were lucky because they were in the nice team and not the mean team. The word "team" was used instead of "group" in line with previous research (Sparks et al., 2017).

After the group manipulation, children's ToM skills towards an ingroup team member or an outgroup team member were tested in a between-group design. Once they completed the five-item ToM scale (Wellman & Liu, 2004) their group based social preferences were tested with the Explicit Preferences task (Qian et al., 2016; Setoh et al., 2019).

Children's between-group differences in performance were expected to be in line with the findings in Study 1 with a few cautionary adjustments. First, it was predicted that performance in the overall scale and specifically in the FB task would be better in the outgroup condition, but the effect would not be as strong as in Study 1. Though minimal groups have been shown to be similarly powerful to real social categories when creating ingroup preferences (Dunham, 2018a), it was predicted that the immediate otherness effect that "race" as a social category produced in Study 1 would not be entirely replicable with the MGP. Whereas race triggers immediate perceptual categorization (Hwang et al., 2021), a MGP would require children to remember the color of their team, adding information to process on top of the task's cognitive load.

Second, the sample of the present study belonged to an individualistic culture (New Haven, US) and consequently the performance of the ingroup condition was expected to follow the sequential order found in individualistic countries DD > DB > KA > FB > RAE (e.g., Wellman & Liu, 2004; Shahaeian et al., 2011). As in Study 1, the ingroup's sequential order could be predicted based on previous experimental evidence because the ingroup condition was tested with a standard scale without a pragmatic manipulation. I did not have a prediction based on evidence for the outgroup condition. The sequence found in Study 1's outgroup did not provide enough guidance because the sample belonged to a collectivist culture and their performance was highly influenced by the boost in the KA task. The heightened performance of the KA task was not predicted for this group.

Third, in the present manipulation character-trait information was added that was not present in Study 1 with "race" as the social category. The MGP allowed me to make the ingroup team "nice" and the outgroup team "mean." By adding character-trait information in the MGP and drawing a competitive difference between the nice team and the mean team, it was expected that the performance difference between conditions in the RAE task would increase. One possibility was that children in the ingroup condition would exhibit a better performance because they would be more likely to project how they would feel onto a similar protagonist (a difference that was not found in Study 1 potentially due to a floor effect). Another possibility could be that a competitive set-up could enhance children's projecting skills. This possibility would be in line with Hackel et al.'s (2014) findings that outgroup threat facilitates outgroup mind perception. In a mind perception paradigm, they found that

participants were more likely to perceive a mind behind an ingroup member, but the pattern changed when the outgroup was presented as threatening.

3.2.2. Method

Participants

I recruited and tested one hundred 4-to-5-year-old children (50 females, M_age= 53 months; SD_age: 6 months) from New Haven preschools (CT, US) and the local museums in the area (The Yale Peabody Museum of Natural History and the Stepping Stones Museum for Children). A power analysis conducted with G*Power (Faul et al., 2009) showed that a sample of 100 participants would have an 80% power to detect a moderate effect size (d=0.5) significant at .05. The racial distribution of the participants was the following: White (65%), Hispanic (5%), Asian (3%), Multiracial (3%) and not specified (24%). The study was approved by the institutional review board of Yale University and a caregiver's informed consent was obtained for each participant.

Measures

MGP assignment

Before the presentation of the tasks, children were assigned to a group through a MGP assignment.

As in Richter et al. (2016), participants were shown two groups of child cartoon characters wearing team-color t-shirts, the orange team, and the green team. See Figure 8. Children were then asked to close their eyes and draw a coin from a bucket. The participant did not see it, but the bucket only contained green coins, so all participants were assigned to the green team. Once the participant opened their eyes, the experimenter showed excitement and said "Congratulations, you are on the green team! Children on the green team are really nice." Then, instead of introducing the orange team in a neutral way like in Richter et al. (2016), the experimenter introduced the orange team in a "threatening" way following the procedure in Nesdale et al. (2005). To manipulate threat the orange team was introduced as follows: "Good thing you are not on the orange team. Children in the orange team are mean."

Figure 8. Recreation of MGP assignment procedure for Study 2.

Children were then asked a manipulation control question "Do you remember which team you are on?" and they were given a green t-shirt to wear throughout the duration of the study.

Theory of mind scale

As in Study 1, participants were tested on the five-item ToM scale (Wellman & Liu, 2004). Instead of pictures of racial ingroup and outgroup protagonists, each task was paired with a gender-matched child cartoon character. See Figure 9. Protagonists were also referred to in a neutral manner (e.g., "this kid") and the only information provided were the teams they belonged to. There were two versions of each task to control for object saliency and the tasks were administered in order of difficulty.

Coding. As in the previous study, children were assigned 1 point if they passed and 0 if they failed. Children's performance was analyzed with the overall score (max.5) as well as by task and Guttman and Rasch scale analyses were applied.
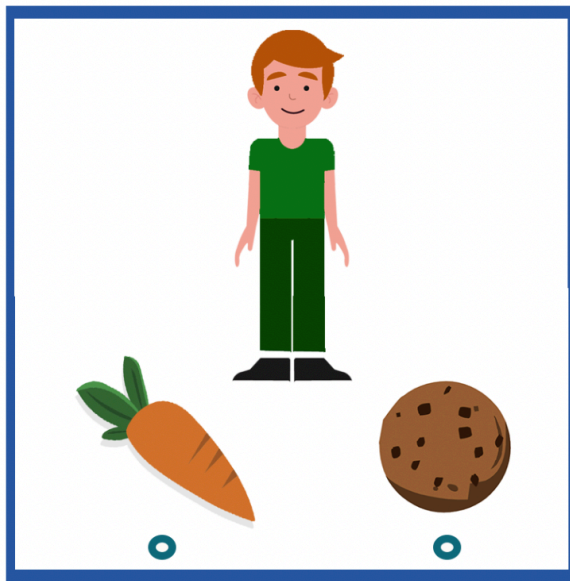
Figure 9. Recreation of the minimal ingroup condition of the DD task for male participants.

Explicit Preferences Task

The same adaptation of the Explicit Preferences task (Qian et al., 2016; Setoh et al., 2019) previously used in Study 1 was administered. Participants had to choose between minimal ingroup team members and outgroup team members in five different scenarios (5 trials). The stimuli depicted child cartoons and was gender-matched to the participant. See Figure 10. The position of the characters was counterbalanced across trials.
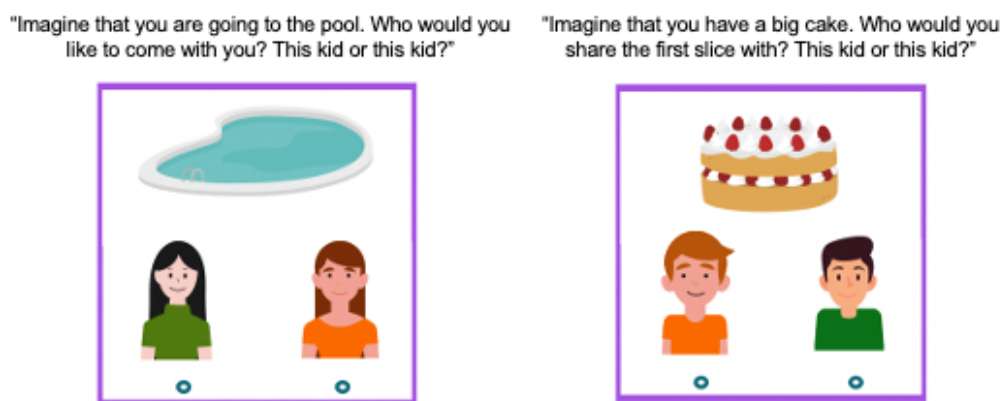
Figure 10. Recreation of a female trial (left) and a male trial (right) of the Explicit Social Preferences Task with MGP cartoon characters.

Coding. They were given one point per trial if they chose an ingroup member and the scores were computed as the proportion of trials were the participant chose an ingroup member.

Procedure

All participants were tested individually by one experimenter during 15-minute sessions in a quiet space in their preschool, in a quiet room at the SCD Lab at Yale, or in a quiet zone of a local museum. Children that were tested in a local museum were recruited on the spot by a research assistant.

After they all had been assigned to the green team, half of the participants (*n*=50) mentalized about an ingroup team member and the other half (*n*=50) mentalized about an outgroup team member. The five-item ToM scale was presented first followed by the Explicit Preferences Task. A Macbook Pro laptop was used to display the materials for the MGP assignment as well as the two test tasks programmed in Qualtrics. Participants were tested in their native language, English.

3.2.3. Results

ToM Scale

It was expected that the demographic sample would be balanced across conditions because condition assignment was random. The sample was balanced in gender (25 female per condition) but when separate *t* tests for age were conducted, a small difference in months that was statistically significant was found; children in the outgroup condition were on average three months younger than their ingroup counterparts, $t(98) = 2.25, p = .026, d = 0.45$. To include age as a variable in the statistical model, *t* tests were conducted instead of regression analyses.

Table 6 shows the percentages of participants passing each task in the ToM scale by condition as well as the ToM overall score. To compare between-group performance, a linear model with the complete five-item score as the dependent variable, group condition as a predictor variable and children's age in months as a control variable (Table 6) was fitted. The model explained a statistically significant and moderate proportion of variance, $R^2 = 0.13$, $F(2, 97) = 7.28, p < .001$, R^2adjusted $= 0.11$. It was found that children's performance in the

overall scale was predicted by group condition, t (97) =2.14, p = .03, as well as by age, t (97) =3.56, p < .001. Children in the outgroup condition (M=2.78) outperformed children in the ingroup condition (M=2.46) in the overall ToM battery.

Table 6. Percentages of participants passing each task in the ToM scale and TOM overall score by condition.

ToM Task	Ingroup condition N =50	Outgroup condition N =50
Diverse Desires	80%	80%
Diverse Beliefs	74%	76%
Knowledge Access	52%	44%
False Belief	22%	32%
Real-Apparent Emotion	18%	46%
ToM five-item overall score	**2.46**	**2.78**

To analyze the performance of each individual task, five separate linear models with each item of the scale as the dependent variable, condition as a predictor, and age as a control variable were ran. Table 7 reports the six linear models (overall scale and by task) and shows that, when analyzed separately, the only task that was significantly predicted by group condition was the Real Apparent Emotion, t(97)=3.13, p < 0.01, where the outgroup performance (M=0.46) was superior to the ingroup's (M=0.18).

Importantly, performance in the FB task was significantly below-chance in both conditions (ingroup: t(49) = -4.73, p < .001, d =-0.67 / outgroup: t(49) = -2.70, p = .009, d =-0.38).

Table 7. Linear model regression for overall and individual task performance with condition as a predictor and age as a control variable.

Predictors	TOM Estimates	CI	p	DD Estimates	CI	p	DB Estimates	CI	p	KA Estimates	CI	p	FB Estimates	CI	p	RAE Estimates	CI	p
(Intercept)	-1.08	-3.08 – 0.92	0.287	0.05	-0.63 – 0.74	0.881	0.00	-0.74 – 0.75	0.994	-0.55	-1.39 – 0.30	0.202	-0.58	-1.33 – 0.18	0.133	-0.01	-0.79 – 0.77	0.979
Condition	0.51	0.04 – 0.98	**0.035**	0.04	-0.12 – 0.20	0.627	0.06	-0.12 – 0.23	0.506	-0.02	-0.22 – 0.18	0.816	0.14	-0.04 – 0.32	0.117	0.29	0.11 – 0.47	**0.002**
Age in months	0.07	0.03 – 0.11	**0.001**	0.01	0.00 – 0.03	**0.030**	0.01	0.00 – 0.03	**0.049**	0.02	0.00 – 0.04	**0.013**	0.02	0.00 – 0.03	**0.037**	0.00	-0.01 – 0.02	0.624
Observations	100			100			100			100			100			100		
R^2 / R^2 adjusted	0.131 / 0.113			0.047 / 0.028			0.040 / 0.020			0.068 / 0.049			0.056 / 0.037			0.092 / 0.074		

Sequential order in the ToM scale

As in Study 1, the individual performance patterns by condition were analyzed first with the Guttman model of scale analysis (Guttman, 1944; 1950). Table 8 shows the patterns found in the present sample. The sequential order found in individualistic cultures captured 62% of the ingroup condition with a coefficient of reproducibility of 0.80 and 50% of the

outgroup responses with a coefficient of reproducibility of 0.74. Similarly to Study 1's Guttman scale analysis, the current data did not conform to a stringent scale.

Table 8. Guttman Scalogram patterns for the five-item ToM scale by condition.

ToM task	Pattern						Other patterns	N
	1	2	3	4	5	6		
Diverse Desires	-	+	+	+	+	+		
Diverse Beliefs	-	-	+	+	+	+		
Knowledge Access	-	-	-	+	+	+		
False Belief	-	-	-	-	+	+		
R-A Emotion	-	-	-	-	-	+		
Ingroup condition	4	7	15	5	0	0	19	50
Outgroup condition	3	1	9	3	4	5	22	50
Total	7	8	24	8	4	5	41	100

Following the procedure in Study 1, the data was then analyzed with the Rasch model analysis (Rasch, 1960) using the eRm package (Mair & Hatzinger, 2007) through R (R Core Team, 2022). Table 9 and Table 10 show the difficulty parameters and item fit statistics for the five items of the ToM scale (ordered from easiest to hardest) for each condition respectively. Both infit and outfit standardized statistics confirmed that the five items in both conditions fit the Rasch model as they fell within between the -2 to 2 range described in Wellman and Liu (2004). Table 9 reveals that the order of item difficulty for children in the ingroup condition replicated the sequence found in individualistic cultures DD > DB > KA > FB > RAE. Table 10 shows that the sequence that resulted from the outgroup's performance deviated from the individualistic or the collectivist sequences. In the outgroup sample, the RAE task moved up two steps in the difficulty scale: DD > DB > RAE > KA > FB.

Table 9. Rasch Model item difficulty levels (easiest to hardest) and fit statistics for the ingroup condition.

ToM Task	Difficulty	SE	Outfit t	Infit t
Diverse Desires	-1.65	0.36	-0.78	-1.29
Diverse Beliefs	-1.24	0.32	1.16	0.67
Knowledge Access	-0.22	0.29	-1.56	-1.97
False Belief	1.42	0.34	-1.27	-1.47
R-A Emotion	1.68	0.36	-0.59	-0.31

Table 10. Rasch Model item difficulty levels (easiest to hardest) and fit statistics for the outgroup condition.

ToM Task	Difficulty	SE	Outfit t	Infit t
Diverse Desires	-1.30	0.35	-1.09	-1.10
Diverse Beliefs	-1.01	0.33	-0.04	-0.60
R-A Emotion	0.50	0.29	1.03	0.69
Knowledge Access	0.60	0.29	-0.56	-0.46
False Belief	1.21	0.31	-0.90	-1.18

Continuing with the procedure in the previous study, the corresponding Item Characteristic Curves were plotted by condition and are displayed in Figure 11. The ICCs confirmed that the ingroup condition followed an individualistic sequential order and the outgroup condition followed a new sequence.
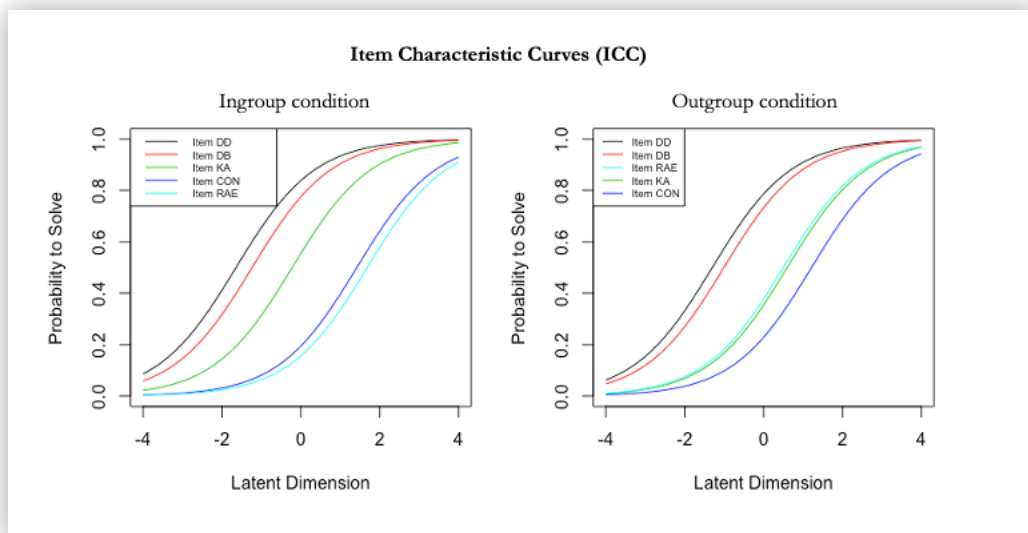
Figure 11. ICC for the ingroup condition (left) and the outgroup condition (right).

ToM and explicit preferences

All children showed an explicit bias favoring their ingroup team members, $t(99)= 25.03$, $p < .001$, $d=2.50$, and the bias did not differ between conditions ($M\_ingroup = 0.92$, $M\_outgroup = 0.91$), $t(97.55)= 0.48$, $p = .63$, $d=0.10$. Moreover, a Pearson correlation analysis showed that explicit bias was not correlated to performance in the ToM scale, $t(98) = 0.40$, $p = .69$, $r =0.04$.

3.2.4. Discussion

The primary goal of this study was to test whether increasing the protagonist's otherness with an artificial category would influence the performance in a FB task. The pragmatic manipulation of the mentalizing target was conducted through a MGP assignment and half of the sample mentalized about an ingroup team member and the other half mentalized about an outgroup team member. To test the MGP effect in children's social preferences their explicit social preferences towards an ingroup and an outgroup were tested. Following the procedure in Study 1, the performance of the FB task was framed within Wellman and Liu's (2004) ToM five-item battery of tasks.

The use of an artificial category instead of a real social category yielded different results than Study 1. As predicted, children's performance in the overall scale was better in the outgroup condition than in the ingroup condition. Surprisingly this boost in performance was not mainly driven by the FB task. Performance in the FB task was descriptively better in the outgroup condition, but it did not reach statistical significance like had been predicted, nor it was strong enough to bring the participants' performance to above-chance levels. In Study 1 it was predicted that an outgroup protagonist would make the protagonist's mental states more salient, would be less knowledgeable and therefore would make the experimenter's question less pragmatically confusing. Performance in Study 1 was superior in the outgroup condition which was interpreted as the result of a potential combination of the three factors that made the task easier for children. A weaker effect had been predicted for Study 2, but a difference was still expected. One possibility is that the "otherness" that was triggered via the minimal group target was not as distinct and salient as the one generated by race. This manipulation was perhaps not strong enough to help children focus on the target's perspective or to activate children's stereotypes on knowledge expectation.

However, another reason that could explain the null finding is that overall performance in the FB task was remarkably below-chance. With such low scores, it could have been that the turnover that was needed from the manipulation was greater than the strength of the manipulation itself.

When tested individually, the only task that yielded a significant difference between groups was the RAE task. Children in the outgroup condition were significantly better. This finding is in line with the second possibility that was presented in the introduction. As in Hackel et al.'s (2014) study, adding a "threatening" character-trait to the outgroup team members

improved children's performance. They were able to project how they would feel and look in the outgroup protagonist's shoes better than in the ingroup ones.

Regarding the sequential order of the scale, the prediction for the ingroup condition was confirmed: children followed the pattern that had been found in individualistic cultures. The sequence that resulted from the outgroup condition's performance was completely new and it was highly conditioned by the RAE task.

Building on the findings from Study 1, the minimal group protagonist manipulation in Study 2 confirmed the pragmatic influence of the protagonist's group membership in the overall ToM scale as well as in the sequential order of passing ToM tasks. This evidence contributes to the pragmatic view on the flexibilization of the competence-performance balance in the development of mentalizing skills.

Chapter 4.

Pragmatic manipulation of the saliency of the protagonist's perspective

4.1. Study 3

4.1.1. Introduction

Study 1 provided evidence that manipulating the protagonist's identity with a social category influenced children's performance in the unexpected-contents FB task as well as in the overall score and sequential passing order of the five-item ToM battery. Children in the outgroup condition outperformed children in the ingroup condition. I hypothesized that the outgroup advantage rested on three pragmatic factors that made the task easier with an outgroup protagonist: the saliency of the protagonist's perspective, the protagonist's expected ignorance and the experimenter's intent. Study 2 showed that a MGP manipulation of the protagonist produced a significant difference in the overall score and sequential passing order of the five-item ToM battery but did not replicate Study 1's outgroup advantage on the FB task. It was suggested that the difference in performance could be due to a weaker "otherness" effect generated by the MGP or due to a potential floor effect that the outgroup manipulation could not overcome.

The goal of Study 3 was to explore the isolated influence of one of the three pragmatic factors of the OAPA: the saliency of the protagonist's perspective. The account predicts that when a task requires mentalizing about a situation in which the participant's perspective and the protagonist's perspective are against one another, presenting an outgroup character increases the weight of the other's perspective and helps children shift from their egocentric stance to the protagonist's perspective. In a way, the greater the distance between oneself and the mentalizing target, the more distinct and salient the target's perspective would be perceived. In this study I wanted to test the pragmatic influence of the saliency of the target's perspective in isolation from the group membership manipulation.

Wellman et al.'s (2001) meta-analysis showed that, in a change-of-location task, when the protagonist's beliefs were pictured (a thought bubble depicting the belief) or stated ("Maxi thinks his chocolate is in the drawer") children's performance improved (without reaching above chance levels). In these examples, the saliency of the target's perspectives was increased in a very overt way. In the present study increasing the saliency of the protagonist's was done by highlighting their need for the object in an expected-contents FB task.

Previous work has shown that a pragmatic manipulation of the protagonist's desire towards the object can yield different results. Using a modified version of a change-of-location FB task (the Duplo task), Rubio-Fernández and Geurts (2016) found that children were more likely to make mistakes in the condition where the protagonist needed the misplaced object, and the object was specifically mentioned in the test question. Upon the reappearance of the protagonist in the adapted change-of-location task, participants were told that the protagonist was hungry and wanted a banana. As reviewed in the introduction, the Duplo task refrains from asking the standard "Where will Sally look for her banana? / Where does Sally think her banana is?" questions, instead it encourages the participant to move the protagonist (a Duplo figurine) and display what will happen next (taking the figurine to the empty cupboard or the cupboard with bananas). Participants in the "hungry" condition failed the FB task because they took the character to the cupboard containing the bananas. The authors showed evidence that increasing the salience of the target object in the test phase directly worsened children's performance in the FB task.

In an unexpected-contents FB task the narrative around the target object is very different than in a change-of-location FB task. The protagonist is only introduced after the content of the box has been revealed and hidden again and the participant is not given any context or narrative about the protagonist. Hence, prior to the test question, the participant has not been tracking the protagonist's perspective as it would happen in a change-of-location task (Rubio-Fernández, 2019). In this pragmatic manipulation I wanted to increase the saliency of the protagonist's perspective by exposing their need for the target object without referencing the target object directly. Previous evidence showed that children performed above-chance in an adaptation of the unexpected-contents task in which they could form a memory representation that associated the protagonist with the expected content of the box (Rubio-Fernández, 2019). It was predicted that the urgency narrative would allow the participant to backtrack a prior relationship between object and protagonist which would increase the participant's focus on the protagonist's perspective.

In a repeated-measures design, children were tested in two conditions (two trials per condition), the "stakes" condition in which the character really needed the object that should have been in the box instead of the unexpected content (Band-Aids in a Band-Aid box) and

the "neutral" condition that followed the set-up of the standard unexpected-contents FB task. It was predicted that children would perform better in the "stakes" condition because the saliency of the protagonist's urgency would provide them with a narrative about the character and their relationship to the object. Focusing on the protagonist's perspective would help them respond the question from a third-person perspective and not from their own situational knowledge.

4.1.2. Method

Participants

Ninety-eight 4-to-5-year-old children participated in the study. They were recruited from the SCD Lab at Yale's database as well as via social media ads and the *Children Helping Science* platform[13]. Twenty-six children were excluded because there was parent/sibling interference during the testing session. The final sample consisted in 72 children (42 female, *M_age*= 58.6 months; *SD_age*: 6.5 months). Children belonged to middle-class families. Racial data was not collected. The study was approved by the institutional review board of the university and a caregiver's informed consent was obtained for each participant.

Measures

Unexpected-contents FB task

Two storylines were created of the unexpected-contents FB task: the neutral condition and the stakes condition (with two scenarios per condition). The neutral condition was a modified version of the original Wellman and Liu's (2004) task that had been used in Studies 1 and 2. In the present version, slight changes were made to the format and the content of the task. First, the design of the task was adapted to an online setting. The child cartoon characters, the settings and the objects were digitally created. The opening of the box and content reveal were depicted through still images presented in sequence.

Second, the check-in question ("Okay, what is in the [Band-Aid box]?") that came right before the test session in the original task was not included. Children's performance in Study

---

[13] https://childrenhelpingscience.com/

2 had been unexpectedly low and, in the current study, the aim was to improve the task's overall performance across conditions. Rubio-Fernández and Geurts (2016) showed that when a check-in question is included right before the test question in a change-of-location FB task, children's performance significantly worsened, even in an open-ended design like the Duplo task. Figure 12 shows the two scenarios that were presented in the neutral condition.

The stakes condition followed the same structure as the neutral condition up to the appearance of the protagonist. Instead of simply presenting the protagonist, children were given information about the protagonist's physical state (was hurt/was hungry) and their need for the object. Figure 13 shows the two scenarios that were part of the stakes condition.
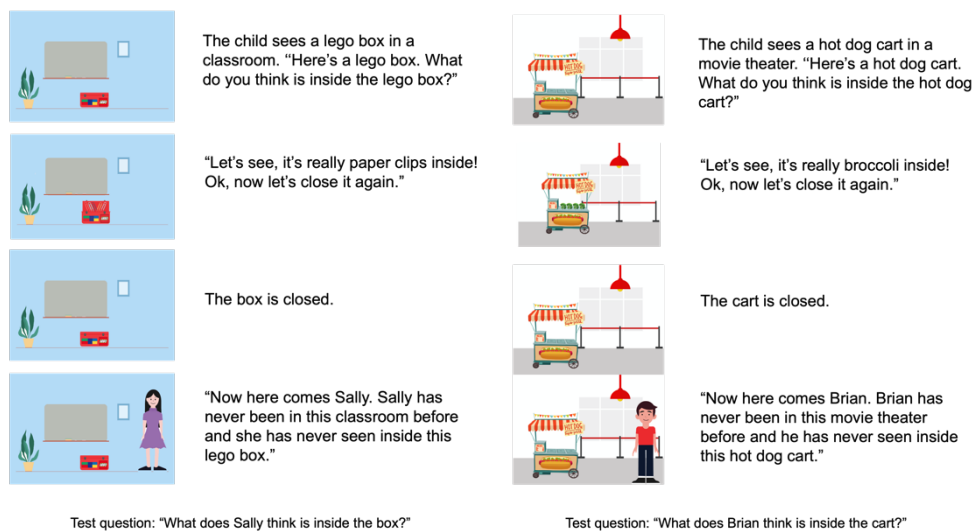
Figure 12. Recreation of the neutral condition of the unexpected-contents FB task: classroom scenario (left) and movie theater scenario (right).

Each child received four unexpected-contents FB trials (two scenarios per condition). The order of the conditions was counterbalanced, and the stimuli was gender-matched to the participants. Distractor images of animals were added between trials and participants were asked easy fun questions about them (e.g., "Do you see the giraffe? Is her neck long or short?").

Coding. Each trial was scored with 1 point if they passed and 0 if they failed. The score for each condition was an average of the two trials.
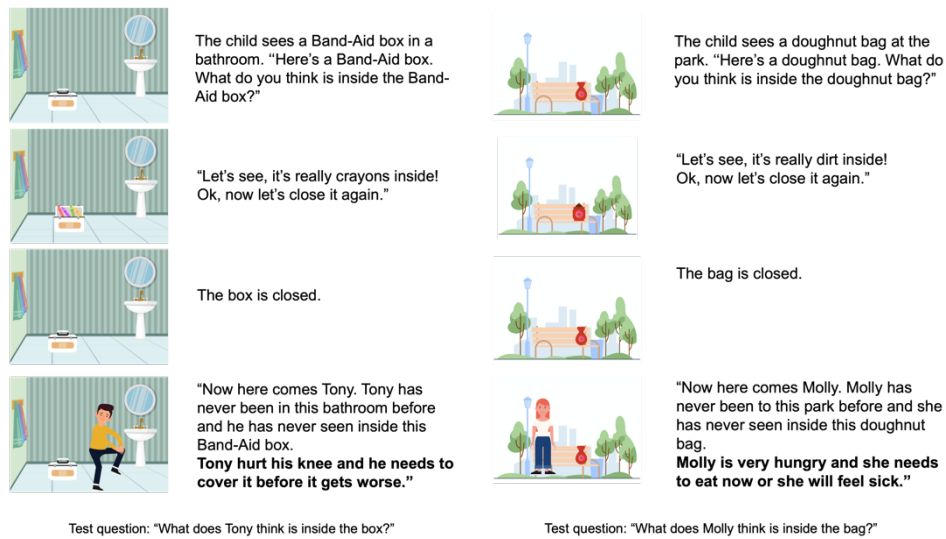
Figure 13. Recreation of the stakes condition of the unexpected-contents FB task: bathroom scenario (left) and park scenario (right).

Procedure

All participants were tested in a moderated online session via Yale's video conferencing platform (Zoom Video Communications Inc., 2016). At the beginning of the session, audio consent was obtained from the parent accompanying the participant. Once the participant had been placed in front of their screen, the parent was asked to step aside and intervene only in the case there was a technical issue. The experimenter then communicated through audio and video streaming with the participant and displayed the stimuli via a shared screen. The test trials were presented in Qualtrics. The experimenter kept navigation control of the survey throughout the testing session. Participants were tested in their native language, English.

4.1.3. Results

Preliminary analyses did not reveal any order effects, therefore, "order" was not included as a factor. The mean score by trial is depicted in Table 11. A repeated measures t-test was carried out to measure the effect that each condition had in the FB score. The test revealed that performance in the stakes condition (M=0.78, SD=0.41) was significantly better than performance in the neutral condition (M=0.67, SD=0.4), $t(143)$= 2.85, p=.005, d=.24. No

performance difference between scenarios in the stakes condition, $t(71)=1.62$, $p=.11$, $d=.19$, or in the neutral condition, $t(71)=0.30$, $p=.77$, $d=.04$ were found.

Table 11. Mean score in the unexpected-contents task by trial and condition.

	FB score
Stake Trials	0.78
Bath	0.82
Park	0.74
Neutral Trials	0.67
Class	0.67
Movies	0.68

4.1.4. Discussion

In this study the aim was to test the saliency of the protagonist's perspective as one of the pragmatic factors that can influence children's performance in an unexpected-contents FB task. The prediction was that highlighting the protagonist's urgency to get the target object would make children perform better at the task because they would be able to mentalize from the protagonist's perspective, which is what the task demands. Furthermore, a protagonist in need would diminish the egocentric biases that might prompt children to answer the question from their own experience.

The present findings confirmed the prediction, children's performance was better in the trials where the storyline emphasized the protagonist's perspective, adding information about their well-being. This finding represents the first evidence of an urgency manipulation in the unexpected-contents FB task and contributes to the evidence found in other FB tasks.

Whereas Rubio-Fernández and Geurts (2016) found that emphasizing the protagonist's hunger and their desire for a banana caused 3-year-olds to fail the change-of-location FB task, in the present manipulation, emphasizing the protagonist's hunger (without explicitly mentioning the object) made children's performance in the unexpected-contents FB task improve. It is important to note that these findings are not in conflict with one another but

instead add evidence to the questionable interchangeability between two tasks that present very different narrative layouts[14].

In Rubio-Fernández and Geurts' (2016) manipulation of the Duplo task, there were two pieces of information provided that were potentially making children fail the task. The first piece of information regarded the protagonist's urgent need related to a physical sensation (i.e., "Lola is very hungry"). The second piece of information was the mention of the target object in the context of need (i.e.,"She wants a banana"). The authors argued that mentioning the target object in this context displaced the focus from the protagonist's perspective to the target object, making it harder for children to pass this task. According to Helming et al.'s (2016) Pragmatic Framework, these findings could also be interpreted under the light of the "cooperative bias." The cooperative bias reflects children's tendency to help others which makes them take a second-person perspective onto the protagonist's instrumental action and provide a "helpful" response. When the stakes of the situation are higher, that is, when the protagonist needs the target object, the motivation to help is ever more present.

The layout of the unexpected-contents task is remarkably different: the object does not belong to the protagonist, nor it has any special meaning to them. Helming et al. (2016) explain that the cooperative bias is irrelevant to account for 3-year-olds' poor performance in this task. In a way, "helping" the agent could imply "warning" the agent of the real content of the box but there is really no guide on how to interpret "warning" in an elicited-response unexpected-contents task. For example, in this manipulation, helping the protagonist could be interpreted as warning them of the real content of the box and saying there are "crayons" in the band-aid box. This interpretation would imply that children's performance would be worse in the stakes condition, however the data revealed the opposite effect.

In the present study, manipulating the protagonist's urgency was used as a prompt to help participants respond from the protagonist's ignorant point of view, by making the protagonist's experience relevant and salient and allowing participants to backtrack a narrative for the protagonist and their relationship to the object. As in Rubio-Fernández and

---

[14] Rubio-Fernández (2019) argues that the two tasks differ in their cognitive demands which can potentially influence performance in different manners. These differences will be reviewed in Study 5.

Geurts (2016) information about the protagonist's physical sensation was provided "Sally is very hungry/Sally's hurt her knee" as well as a need for the target object "She needs to eat immediately/She needs to cover it immediately". According to the predictions for this manipulation, children were better in the stakes condition because they were able to focus on the protagonist's perspective and their relation to the object. Alternatively, their relation to the object could have highlighted the original object itself, leading children to answer the question correctly.

Leaving aside the manipulation of the protagonist's needs, children's overall performance in the present study was significantly better than children's performance in the unexpected-contents FB task in Study 2. This difference was attributed to the two changes that were made in this study from Wellman and Liu's (2004) original task: the removal of the control question and the adaptation to an online setting. On the one hand, removing the control question might have reduced the saliency of the wrong object further enabling children's focus on the protagonist's perspective (Rubio-Fernández & Geurts, 2013).

On the other hand, adapting the task to an online setting removed the intrinsic physicality of the wrong object in the original task, which could have made the task easier. In the in-person testing session of the unexpected-contents FB task the child saw the content of the box and, once the box was closed, the unexpected content was still physically there. Thus, the wrong object was more salient than the object that should have been there because one was present in the room and the other was not. Rubio-Fernández and Geurts (2016) explained that previous research (Freeman & Lacohée, 1995; Mitchell & Lacohée, 1991) has shown that having the two objects in the task improves children's performance because it reduces the saliency of the physicality of the wrong object.

In this study, seeing a digital cartoon of a Band-Aid box that contained crayons might have been less powerful than seeing a real-life Band-aid box. Once the cartoon box was closed, the image of the unexpected content disappeared and the perceptual link to the unexpected content was broken. Thus, the online adaptation might have made inhibiting the unexpected content of the box easier. This interpretation would be in line with Wellman et al.'s (2001) findings that the presence of the object in the task was one of the four variables that influenced performance.

4.2. Study 4

4.2.1. Introduction

Study 3 provided evidence that the saliency of the protagonist's perspective is indeed a pragmatic factor that can enhance performance. The goal of the current study was to examine the pragmatic influence in performance of combining two saliency boosters: 1) increased saliency of the protagonist's perspective through the presentation of the protagonist's needs for the object and 2) the increased saliency of the protagonist's perspective that is inherent to an outgroup target.

The task in this study integrated Study 2's MGP assignment paradigm and Study 3's stakes manipulation. In a between-group design, 4-to-5-year-olds were assigned to one of four conditions: ingroup neutral condition, ingroup stakes condition, outgroup neutral condition, and outgroup stakes condition.

Two possible interactions between the two salience boosters (stakes and outgroup) were contemplated. The first possibility was that the otherness effect would be countered by the urgency effect. That is, an ingroup target's perspective in a stakes situation would be significantly more salient than an outgroup's target perspective in the same stakes situation. According to this possibility, the ingroup stakes condition (+1) would present a similar performance to the outgroup neutral condition (+1). The outgroup stakes condition (+1-1) and the ingroup neutral condition (0) would yield a lower score. This possibility would only be plausible under a cooperative interpretation of the task: the participant would be more likely to want to help an ingroup target than an outgroup target. Considering that the unexpected-contents task does not really allow for a helpfulness interpretation this possibility was discarded.

An additive interaction was predicted instead. This possibility predicted that the effect of urgency (+1)[15] would be added to the otherness effect (+1), creating an even stronger advantage to pass the task (+2). If this were true, the outgroup stakes condition (+2) would yield the best performance: a protagonist that was in need would keep children's focus in the

---

[15] (+1) refers to a booster. The two between-group conditions with one saliency booster are scored (+1), the condition with the two saliency boosters is scored (+2) and the condition with no saliency boosters is scored (+0).

target's perspective and the target's perspective would already be more salient for being an outgroup protagonist. Consequently, the outgroup neutral condition (+1) and the ingroup stakes condition (+1), each having one of the two salience boosters, would yield a similar performance. Lastly, the condition in which there were no boosters, the ingroup neutral condition (+0), would present the worse performance.

In addition to the FB trials (each participant received two trials) a TB trial was included as a control measure for children's performance. As explained in the general introduction, the TB trial follows the same structure as a FB trial, but the protagonist is present during the box reveal and thus holds a veridical belief about the content of the box. Extensive research has shown that when children start passing the FB task, they start failing the TB task (Oktay-Gür & Rakoczy, 2017; Rakoczy & Oktay-Gür 2020; Schidelko et al. 2022; Wellman et al., 2001). Here I wanted to explore if the influence of group membership and urgency in children's FB performance inversely correlated with children's performance in the TB trial.

Following the procedure in Study 2, ingroup preferences based on the MGP manipulation were tested with the Explicit Preferences Task (Qian et al., 2016, Setoh et al., 2019).

4.2.2. Method

Participants

Ninety-six 4-to-5-year-old children participated in the study[16]. They were recruited from the SCD Lab at Yale's database as well as via social media ads and the *Children Helping Science* platform. Sixteen children were excluded because they were not able to complete the assignment. The final sample consisted in 80 children (39 female, *M_age*= 52 months; *SD_age*: 8 months) that were randomly assigned into four conditions (20 children per condition). Children belonged to middle-class families. Racial data was not collected. The study was approved by the institutional review board of the university and a caregiver's informed consent was obtained for each participant.

---

[16] Our recruiting goal was to have 20 participants per condition. The online testing experience in Study 3 showed us that we would have to recruit a larger sample to be able to include valid performance data of 80 participants.

Measures

MGP assignment

All participants were assigned to the green team with the online adaptation of the MGP assignment previously used in Study 2. In the present study I wanted to create a cohesive online environment in which children could easily understand how an ingroup or an outgroup team member would appear in the scene. For this reason, the trials were set up in a school scenario as shown in Figure 14. Instead of drawing a colored coin from a bucket, children were asked to digitally flip a coin to see which group they would be on.

"Welcome! Today we will play a game at a new school. In this school there will be two teams, the green team and the orange team."

"The kids on the green team are very nice."

"And the kids on the orange team are very mean."

"Let's flip a coin to see which team you'll be on!"

"Yay, you're on the green team, they're very nice! Good thing you are not on the orange team because they are very mean."

Control question: Now, before we start, do you remember which team you are on?

Figure 14. Recreation of Study 4's MGP assignment procedure.

Unexpected-contents False Belief Task

As in Study 3, there were two storylines of the unexpected-contents FB task (the neutral trial and the stakes trial) and, in this study, two versions of the protagonist's group membership (ingroup target and outgroup target) were added. This resulted in four conditions: (a) ingroup neutral, (b) ingroup stakes, (c) outgroup neutral, (d) outgroup stakes. Two scenarios per condition were created and counterbalanced across participants, and the stimuli was gender-matched to the participant. Figure 15 shows four trial examples out of the sixteen different trials (2_{target} x 2_{stakes} x $2_{scenarios}$ x 2_{gender}) that were generated for the study.

Coding. Each child received two unexpected-contents FB trials. Each trial was scored with 1 point if they passed and 0 if they failed. The score for each condition was an average of the two trials.

The child sees a lego box in a classroom. "Look here's a classroom. And there's a lego box. What do you think is inside the lego box?"

"Let's see, it's really paper clips inside! Ok, now let's close it again."

The box is closed.

"Now here comes Mary, she is on the green team. Mary has never been in this classroom before and she has never seen inside this lego box."

Test question: "What does Mary think is inside the box?"

The child sees a sidewalk chalk box in a classroom. "Look here's a classroom. And there's a chalk box. What do you think is inside the chalk box?"

"Let's see, it's really puzzle pieces inside! Ok, now let's close it again."

The box is closed.

"Now here comes Mike, he is on the orange team. Mike has never been in this classroom before and he has never seen inside this chalk box."

Test question: "What does Mike think is inside the box?"

The child sees a Band-Aid box in a bathroom. "Look, here's a bathroom! And there's a Band-Aid box. What do you think is inside the Band-Aid box?"

"Let's see, it's really crayons inside! Ok, now let's close it again."

The box is closed.

"Now here comes John, he is on the green team. John has never been in this bathroom before and he has never seen inside this Band-Aid box. **John hurt her knee and he needs to cover it before it gets worse."**

Test question: "What does John think is inside the box?"

The child sees a lego box in a classroom. "Look here's a classroom. And there's a pizza box. What do you think is inside the pizza box?"

"Let's see, it's really rocks inside! Ok, now let's close it again."

The box is closed.

"Now here comes Jen, she is on the orange team. Jen has never been in this classroom before and she has never seen inside this pizza box. **Jen is very hungry and she needs to eat something now or she will feel sick."**

Test question: "What does Jen think is inside the box?"

Figure 15. Recreation of the female ingroup neutral condition trial (top left), male outgroup neutral condition trial (top right), male ingroup stakes condition trial (bottom left) and female outgroup stakes condition trial (bottom right).

True Belief Task

The TB task followed the same structure as the FB task, the difference was that the protagonist was present during the box reveal. An ingroup version and an outgroup version were created. Stimuli was gender-matched to the participant. Figure 16 depicts two examples of the TB trials. Each participant received only one TB trial once they had completed the two FB trials. *Coding.* Trials were scored with 1 point if they passed and 0 if they failed.
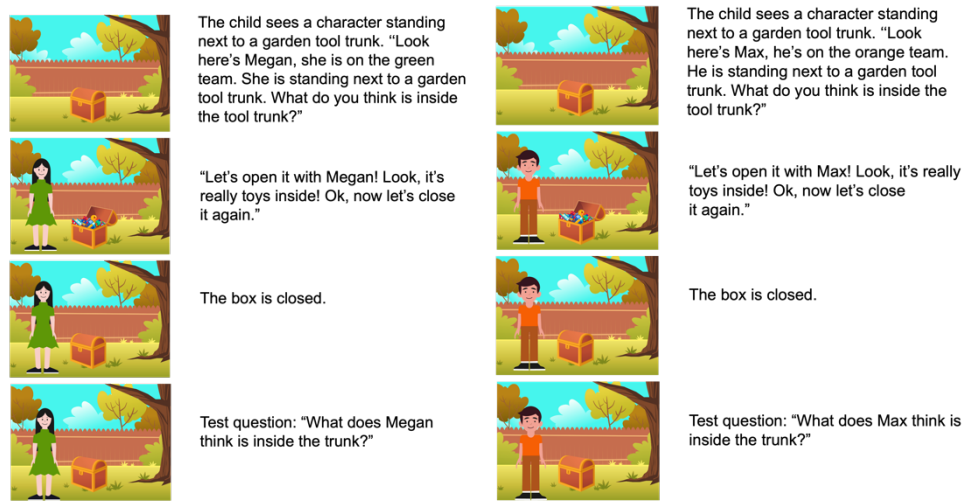
Figure 16. Recreation of the female ingroup TB trial (left) and the male outgroup TB trial (right).

Explicit Preferences Task

The stimuli of the Explicit Preferences task (Qian et al., 2016; Setoh et al., 2019) was the same that was used in Study 2. Participants had to choose between minimal group ingroup team members and outgroup team members in five different scenarios (5 trials). Stimuli was gender-matched to the participant and the position of the characters was counterbalanced across trials. *Coding.* They were given one point per trial if they chose an ingroup member and the scores were computed as the proportion of trials were the participant chose an ingroup member.

Procedure

Individual participant testing was conducted in a moderated online session via video conferencing platform (Zoom). The experimenter first interacted with the parents of the participant to obtain audio consent before the testing session. Parents were asked to be present throughout the study and only intercede if the child needed their technical assistance. Experimenters communicated with participants through audio and video streaming and presented the stimuli via a shared screen.

First children were placed in a team through the MGP assignment. Then, participants were randomly assigned to one of four conditions (ingroup neutral, ingroup stakes, outgroup neutral, outgroup stakes) and completed two FB trials and one TB trial. Finally, children's ingroup bias was tested with the Explicit Preferences Task. The tasks were programmed in Qualtrics. The testing sessions lasted between 15 minutes and 25 minutes, depending on the child's concentration and willingness to participate. The experimenter kept navigation control of the survey throughout the testing session. Participants were tested in their native language, English.

4.2.3. Results

Unexpected-contents FB

The mean scores by condition are shown in Table 12. A 2(Protagonist's group membership: Ingroup/Outgroup) x 2 (Task Type: Neutral/Stakes) ANOVA revealed a main effect of the protagonist's group membership, $F(1,76) = 11.78$, $p < .001$, $\eta_p^2 = .13$. Tukey HSD post-hoc comparisons showed that performance in the ingroup neutral condition ($M=0.45$, $SD=0.44$) was significantly worse than performance in the outgroup neutral condition ($M=0.86$, $SD=0.28$, $p=.003$) and in the outgroup stakes condition ($M=0.83$, $SD=0.34$, $p=.012$).

Table 12. Mean trial scores for the Unexpected-contents FB & TB tasks by condition ($n=20$ per condition).

	Unexpected-contents FB task	TB Task
Ingroup Neutral Condition	0.45	0.75
Ingroup Stakes Condition	0.68	0.85
Outgroup Neutral Condition	0.88	0.7
Outgroup Stakes Condition	0.83	0.5

Moreover, one-sample *t*-tests revealed that children's performance in the FB task was above-chance in both neutral and stakes outgroup condition (neutral: $t(19) = 6.10$, $p < .001$, $d=1.36$; stakes: $t(19) = 4.33$, $p < .001$, $d=0.97$). Children's performance in the ingroup conditions was not different from chance (neutral: $t(19) = 0.52$, $p = .60$, $d=0.12$; stakes: $t(19) = 1.79$, $p = .09$, $d=0.40$).

TB task

Regarding children's performance in the TB task, one-sample *t*-tests showed that performance was above-chance for both ingroup conditions (neutral: $t(19) = 2.52$, $p = .021$, $d=0.56$; stakes: $t(19) = 4.27$, $p < .001$, $d=0.96$) and at-chance in both outgroup conditions (neutral: $t(19) = 1.90$, $p = .072$, $d=0.43$; stakes: $t(19) = 0.00$, $p > .999$, $d=0$). The only condition in which performance in the TB task was significantly different than performance in the FB task was in the outgroup stakes condition, $t(19) = 2.94$, $p = .008$, $d=0.66$.

FB and explicit preferences

All children showed an explicit bias favoring their ingroup team members, $t(79)= 22.22$, $p < .001$, $d=2.48$, and the bias did not differ between group membership conditions ($M\_ingroup = 0.91$, $M\_outgroup = 0.92$), $t(66.63)= 0.26$, $p = .78$, $d=0.06$. Moreover, a Pearson correlation analysis showed that explicit bias was not correlated to performance in the FB tasks, $t(78) = 0.31$, $p = .76$, $r =0.04$.

4.2.4. Discussion

The main objective of the present study was to examine children's performance in the unexpected-contents FB task combining two manipulations that attempted to increase the saliency of the protagonist's perspective: the need for the object (urgency effect) and the group membership of the target (otherness effect). Two potential interactions between the manipulations were contemplated: a counter effect and an additive effect. The counter effect would imply that the ingroup stakes condition would yield a similar performance to the outgroup neutral condition and the outgroup stakes condition would yield a similar lower performance to the ingroup neutral condition. It was predicted that the interaction would be additive: the outgroup stakes condition would have the best performance, followed by the

ingroup stakes condition and outgroup neutral condition, and leaving the ingroup neutral condition in the last place.

The data revealed that, as could be explained by both possibilities, children in the ingroup neutral condition yielded the worst performance. It also showed that performance in the ingroup stakes condition and the neutral outgroup condition was not statistically different (this performance pattern was also predicted by both possibilities). The data point that brought the data closer to my hypothesis (an additive effect of the two boosters) was the better performance of the outgroup stakes condition in relation to that of the ingroup neutral condition. However, the prediction could not be fully confirmed because the outgroup stakes condition did not significantly differ from the outgroup neutral condition or the ingroup stakes condition.

The present findings added evidence to the outgroup protagonist booster effect that had been found in Study 1 and only descriptively in Study 2: children in the outgroup conditions performed significantly better than children in the ingroup neutral condition. Moreover, while performance in both ingroup conditions was at-chance, performance in the outgroup conditions reached above-chance levels. Regarding the urgency booster effect, children in the ingroup stakes condition performed descriptively better than children in the ingroup neutral condition, but the difference did not reach statistical significance. An urgency booster effect in the outgroup conditions was not found. This could be explained by the overall high scoring in the outgroup conditions. The unexpected-contents FB task only provides binary data and therefore the measurement might not be nuanced enough to distill between performances that are already almost proficient. The absence of the urgency booster effect in the outgroup conditions also provides support that the effect of the urgency manipulation in Study 3 came from helping the participant focus on the protagonist's perspective and their relation to the object, and not just from the fact that presenting a need for the object highlighted the object itself.

Interestingly, the passing pattern of TB trials was negatively correlated with the passing pattern in FB trials. Children in the ingroup conditions performed above-chance in the TB trials (compared to the at-chance in FB trials) whereas performance in the outgroup conditions was not different from chance in TB trials (compared to the above-chance in FB

trials). Further, the only condition in which TB-FB performance significantly differed was the outgroup stakes condition, which had been predicted that it would exhibit the highest FB performance.

These results could be interpreted under the light of the third pragmatic factor of the OAPA: the protagonist's expected ignorance. Based on Ames' (2004) double inferential mechanism for mind reading account (projection from self-template versus stereotyping) and Westra's (2019) account on mindreading and stereotypes, this pragmatic factor relies on children's applying a knowledge-related stereotype to the outgroup protagonist: the outgroup protagonist does not share my privileged knowledge, and the outgroup protagonist is generally less competent. Since a knowledgeable target makes the task harder (i.e., adults – Seehagen et al., 2018; competent targets – Zmyj and Seehagen, 2020; a supernatural being – Lane et al., 2010), the expected ignorance of the outgroup target should improve children's performance.

In the ingroup neutral condition, the participant would expect the protagonist to share their knowledge as an ingroup teammate. Projecting the participant's knowledge onto the target would lead to an incorrect answer in the FB task and a correct answer in the TB task. On the opposite side, in the outgroup stakes condition, a condition in which the saliency of the protagonist's perspective was boosted through both the urgency effect and the otherness effect, the participant would not expect the protagonist to share their real-world situational knowledge and would stereotype them as less knowledgeable in general. Expecting the protagonist's ignorance would lead to a correct answer in the FB task and an incorrect answer in the TB task.

Study 5 was designed to further analyze the booster effect of the protagonist's expected ignorance/knowledge by using a paradigm in which there were two agents that could either share or not share group membership.

Chapter 5.
Pragmatic manipulation of the protagonist's expected ignorance

5.1. Study 5

5.1.1. Introduction

The experimental effort of the four previous studies relied on the performance in one of the two standard FB tasks: the unexpected-contents FB task. Though for the most part this task has been treated as an equivalent of the change-of-location task, there are notable differences between the respective set-ups of the tasks that question the assumption of interchangeability between them. Rubio-Fernández (2019) points to three aspects that differ between these tasks. One aspect was already discussed in Study 3: the narrative of the task. In the change-of-location task the participant is presented with a narrative about the protagonist that allows the participant to track the protagonist's perspective until the test phase. In the unexpected-contents task, such narrative is completely absent; the protagonist is introduced without any context at the same time the test question is asked. The second aspect is closely connected to the first one. In a change-of-location task, the narration allows children to form a memory trace that associates the protagonist with the object in the first location. Such association is not possible in the unexpected-contents task.

A third aspect regards to the difference in the physicality of the correct response in the two FB tasks. Rubio-Fernández (2019) points out that, whereas the two possible responses are depicted in the test phase of the change-of-location task, only the wrong response is physically present in the unexpected-contents task. The physicality of the wrong object in the unexpected-contents task was addressed in Study 3's discussion when comparing children's better performance in an online paradigm (no perceptual link to the object during the test phase) versus an in-person paradigm (the wrong content is physically present in the room). The difference in performance was attributed to the effect of the physical presence of the object, which has been observed as an influential variable in performance (Wellman et al., 2001). Interestingly, Schidelko et al. (2021) found that online testing yielded the same results as lab testing with a change-of-location task. This finding provides further support to the "physicality" differential aspect between the two tasks.

There is yet a fourth aspect that differs between tasks: the number of agents featured. The unexpected-contents task features one character, the protagonist or mentalizing target. As

seen in the previous studies, testing the influence of the protagonist's identity as a pragmatic performance factor in this task implies simply manipulating the identity of the mentalizing target and making it either an ingroup or an outgroup to the participant. Contrastingly, the change-of-location task presents two agents, the protagonist and the agent that moves the object. The set-up of the task establishes three inter-subject dyads: protagonist-agent, agent-participant, and protagonist-participant.

In this study the identity of the protagonist and the agent were manipulated, and I explored the pragmatic influence on performance that a shared-group membership between protagonist-agent and protagonist-participant would have on the protagonist's expected knowledge/ignorance (OAPA's second pragmatic factor). The MGP assignment from Studies 2 and 4 was repeated and children were assigned to one of four between-group conditions: the ingroup-ingroup condition (shared-group membership between protagonist-agent-participant), the ingroup-outgroup condition (shared-group protagonist-participant dyad), the outgroup-ingroup condition (shared-group agent-participant dyad) and the outgroup-outgroup condition (shared-group protagonist-agent dyad).

Instead of using the standard format of the change-of-location task that provides dichotomic pass/fail responses I opted for a format that would be able to detect nuance between the conditions: the Sandbox task. This paradigm was created by Huttenlocher et al. (1994) and was later adapted by Bernstein et al. (2011). In recent years it has been used by many researchers (Sommerville et al., 2013; Coburn et al., 2015; Haskaraka Kizilay et al., 2022; Ni et al., 2022; Samuel et al., 2018a; 2018b; Speiger et al., 2021; 2022).

In the adapted format of the change-of-location task, the participant sees the protagonist and the agent standing on top of a horizontal box (a sandbox). The protagonist places the object in the sand (a non-descript discrete location along the horizontal box – Location A), the object disappears in it and the protagonist leaves. During their absence, the agent removes the object from Location A and buries it at another non-descript discrete location in the sandbox (Location B). When the protagonist comes back the participant is asked to point to the location where the protagonist thinks the object is. Unlike in the standard FB tasks, in this task the object's location is not limited to two containers and thus it allows for a metric measurement (in cm) of the participant's response on the location of the object.

In the present study a child-friendly version of Samuel et al.'s (2018a) online adaptation of the task[17] was built. Children completed the trials through moderated online sessions and gave their response by clicking on their screen with the mouse or trackpad. Children's bias scores were calculated by taking the difference in pixels between the correct location and the location the participant clicked on and dividing it by the length of the sandbox. A positive bias score (significantly different than zero) would mean that children chose a location closer to the wrong location than the correct one. Pilot testing showed that 4-year-olds struggled with the task demands in the remote online format, so 5-6-year-olds were recruited instead.

The main research goal was to test the OAPA's predictions in a standard change-of-location FB task adapted to the Sandbox format in which both agents' group memberships were manipulated. To be able to compare the effect with other paradigms in which the protagonist's perspective differed (or not differed) from the participant's, three additional within-subject trials were included: an altercentric trial, a TB trial, and a prank trial.

The standard FB trial followed the narrative of a regular change-of-location task. It was expected that the group membership manipulation would increase children's bias towards the wrong location in three conditions: the condition in which the protagonist, the agent, and the participant shared group membership (ingroup-ingroup), the condition in which the protagonist and the agent shared group membership (outgroup-outgroup), and the condition in which the protagonist and the participant shared group membership (ingroup-outgroup). It was hypothesized that the shared-group membership would bring the participant to expect a shared knowledge between protagonist-agent / protagonist- participant that would bias their response towards the wrong location. It was predicted that the protagonist's expected knowledge that resulted from a shared-group membership would operate as a pragmatic factor that would worsen children's performance. In the outgroup-outgroup condition the participant would expect the protagonist's knowledge to be in synch with the agent's knowledge. In the ingroup-ingroup condition the participant would expect the protagonist's

_____

[17] We adapted Samuel et al. (2018a)'s Sandbox *format* to a child-friendly version. That is, we used the exact sandbox measurements (length of sandbox, x-coordinates for Location A and Location B) and bias calculations that appeared in Samuel et al. (2018a). However, we were interested in different experimental questions and thus the *content* of the task (the trial types) was different.

knowledge to be in synch with the agent's knowledge and their own knowledge. Lastly, in the ingroup-outgroup condition the participant would expect their ingroup protagonist to share their knowledge.

The altercentric trial followed the same structure as the FB trial up to the test question when instead of asking about the protagonist's false belief, the participant was asked about the current location of the object (i.e., "Where is the car now?"). This trial was included to test whether the presence of an ignorant ingroup or outgroup protagonist in the test question would interfere with the participant's own perspective and bias them to locate the object closer to the location where the protagonist left it.

Evidence for the existence (or the non-existence) of an altercentric bias in children and adults' mindreading processes has been mostly provided from visual perspective-taking studies. Samson et al. (2010) first reported an altercentric intrusion effect in adult's visual perspective-taking with the "dot perspective task". In this task the participant sees a room with two walls in front of each other and a human avatar in the middle facing one of the two walls. On each wall there are dots, and the participant is asked to report how many dots there are from their own perspective or from the avatar's perspective. The researchers referred to an altercentric intrusion effect to explain why participant's response times were slower and less accurate when they had to report the dots from their own perspective and their perspective differed from the avatar's view. This effect was interpreted under the lens of Apperly and Butterfill's (2009) Two-Systems Account of mindreading. That is, the interference resulted from System's 1 flexible and fast mindreading cognitive processes that computed the human avatar's line of gaze and were resistant to strategic control. Replication efforts in subsequent research fell into opposite directions. Rubio-Fernández et al. (2022) explains that some replication efforts provided confirmative evidence and enriched an automaticity interpretation in visual perspective taking as part of a set of theory of mind skills (Simpson & Todd, 2017) while others provided evidence that the interference effect could also happen with inanimate objects (an arrow) pointing to general-domain cognitive processes such as attentional orientation (Santiesteban et al., 2014). More recently, Vestner et al. (2022) found that depicting desk fans ("a class of inanimate object known to cue attention", p.1) produced a comparable altercentric effect (i.e., slower in performance in inconsistent trials) than a human target. Lastly, in one of the latest uses of the dot task (a

simplified version) the authors did not find an altercentric interference in children or adults until they incremented the Executive control demands. They attributed the altercentric interference to the perspective-switching costs of the paradigm instead of the result of an automatic mindreading process that is activated by the presence of an avatar (Rubio-Fernández et al. 2022).

Aside from the visual perspective-taking research studies, recently, a few research groups have tested the altercentric bias in change-of-location paradigms using the Sandbox task. Haskaraka Kizilay et al. (2022) found no evidence of bias in adults in either altercentric or egocentric trials (egocentric trials would be the equivalent to the standard FB sandbox trial). Speiger et al. (2022) found that an altercentric bias only appeared when the adult participants had previously answered an egocentric trial. Lastly, Ni et al. (2022) reported no evidence of an altercentric bias in 3-year-olds.

Considering the mixed evidence of an altercentric bias in visual perspective-taking and the weak evidence for an altercentric bias in a Sandbox task, first and foremost I wanted to test the bias itself. The between-group predictions mirrored those of the standard FB trial, shared group-membership would bias the participant to the wrong location of the object.

The TB trial was included to further explore the negative correlation found in Study 4, where participants in the ingroup conditions passed the FB task but were at chance in the TB task and participants in the outgroup condition followed the inverse pattern. According to past research on the correlation between FB and TB performance (Oktay-Gür & Rakoczy, 2017; Rakoczy & Oktay-Gür, 2020; Schidelko et al., 2022; Wellman et al., 2001), children in this study (5-to-6-year-olds) would fall in the age range where they would be close to mastering the standard formatted change-of-location task and fail the standard-formatted TB task. It was expected that children in the four conditions would exhibit a larger bias in the TB trial versus the FB trial. In line with the predictions for the two previous trials, it was expected that children in both shared-group conditions and the ingroup-outgroup condition would exhibit a larger bias than children in the outgroup-ingroup condition.

The prank trial was included as a control measure for a proficient performance. In this trial the protagonist buried the object in the grass and left. While the protagonist was away, an

external force (the wind)[18] moved the object from Location A to Location B. Then the prank element was introduced as follows: "Let's play a funny trick on Sam! Let's hide a rubber bug where he thinks the [object] is! Where should we hide the rubber bug?" This trial differed from the other three trials in two aspects: 1) there was only one agent featured (the protagonist) as the object was moved by an external force and, 2) the protagonist was not present in the test slide. In the initial mock-up of the task, the protagonist was present in the test slide. Pilot testing showed that having the protagonist present made children place the bug right at the protagonist's feet. Once the protagonist was removed from the test slide children were able to give an answer to the test question and point to one of the two locations. It was expected that this trial would be the easiest (with the least positive bias scores) because two of the four variables that influenced performance according to Wellman et al.'s (2001) meta-analysis ("motive" and "participation") were manipulated. The action in the prank trial was framed in a trick narrative ("deceptive motive") and the participant was substantially involved in the deception ("participation"). The two aspects were manipulated during the test-phase in the prank trial, instead of the set-up phase (as described in Wellman et al., 2001). It was expected that children would be more biased in the ingroup conditions, but it was predicted that performance would be very accurate without much room for improvement.

In Wellman et al.'s (2001) meta-analysis, the authors found that two of the task manipulations that improved performance were framing the task in terms of trickery and asking the child to participate in the deception during the set-up. In the prank trial the trickery framing and the participant's involvement in the deception took place in the test phase instead of in the set-up of the task. This variation could have enhanced even more the trickster boost effect in the prank trial.

Previous studies that used the Sandbox task included distractors between tasks to prevent participants from using perceptual strategies and fixating their gaze on the first location[19]. In the study four different types of trials were combined and each type had a different

---

[18] The initial design depicted an arrow that moved the object and the experimenter said that "someone" moved it. Using the wind as the external force that moved the object was a much better solution that we took from Speiger et al. (2021).

[19] For example, Bernstein et al. (2011) used a "Where's Waldo" search distractor. Coburn et al. (2015) and Samuel et al. (2018a; 2018b) used a word search puzzle.

instruction and test question, therefore there was no need to add distractors to block visual fixation. Nonetheless, due to the high processing demand of the task and to prevent perspective-switching costs, a second task was included to be alternated with the Sandbox task: the Social Activity FB task created by Nguyen and Frye (1999). Structured very similarly to a change-of-location paradigm, in this task there are two agents that are conducting an activity together. Then the protagonist leaves the scene and while they are gone the agent switches the activity for another one. The test question is "What does the protagonist think the agent is doing in the room? [Activity A] or [Activity B]?" In Nguyen and Frye's (1999) study, 5-year-olds had more trouble passing the FB social task than a regular change-of-location task.

The procedural goal was to alternate each Sandbox trial with a Social Activity FB trial. Since four trial types had been created for the Sandbox task, I took the opportunity to also create four different versions of the Social Activity FB task to explore how children's performance could be influenced with a group-membership manipulation. Four trial types were created combining two factors: the number of agents and the nature of the social activity. Regarding the number of agents, I made two individual trials that followed the original design with two agents (the protagonist and the agent that switches activity) and two group trials in which the agent was replaced with a group of children. Regarding the nature of the social activity, I created two trials in which Activity B was a neutral activity very similar to Activity A (e.g., playing soccer vs. playing basketball) and two trials in which Activity B was a mischievous activity. The Social Activity FB task, like the original change-of-location task, provides binary results and thus is not very sensitive to subtleties in performance. Since the participant population fell in the age range that should already master FB attribution in forced-choice paradigms, I was cognizant about the possibility that no between or within group differences would be detected. Like in the FB Sandbox trial, it was predicted that children in the shared-group conditions would be more likely to struggle with the task. It was also predicted that the group trials would be easier than the individual trials and that the mischievous trials would be easier than the neutral trials.

5.1.2. Method

Participants

Eighty-five 5-to-6-year-old children participated in the study[20]. They were recruited from the SCD Lab at Yale's database, via social media ads, the *Children Helping Science* platform, and at recreational parks in the New Haven area. Twenty-one children were excluded because they were not able to complete the assignment. The final sample consisted in 64 children (26 female, M_age= 72 months; SD_age: 7 months) that were randomly assigned into four conditions (16 children per condition). Children belonged to middle-class families. Racial data was not collected. The study was approved by the institutional review board of the university and a caregiver's informed consent was obtained for each participant.

Measures

MGP assignment

All participants were assigned to the green team following a similar procedure than in

Study 4 with two differences: it was not set-up in a school, and the team images featured male and female cartoon characters. See Figure 17.

"Welcome to this game! In this game there are two teams, the green team and the orange team. The kids on the green team are very nice..."

"And the kids on the orange team are very mean!"

"Let's flip a coin to see which team you'll be on! Tap on the coin, then hit next."

"Yay, you're on the green team, they're very nice! Good thing you are not on the orange team because they are very mean."

Control question: Now, before we start, do you remember which team you are on?

Figure 17. Recreation of the MGP assignment for Study 5.

[20] Our recruiting goal was to have 16 participants per condition. The online testing experience in Study 3 and 4 showed us that we would have to recruit a larger sample to be able to include valid performance data of 64 participants.

The Sandbox task

A child-friendly version of the task was created following the exact measurement guidelines of Samuel et al.'s (2018a) online adaptation of the pen and paper version by Sommerville et al. (2013). In Samuel et al.'s (2018a) participants saw a slide with the text describing a scene on top and a rectangular box below. The protagonist and the agent were only referred to by name without any visual aid. The objects varied between trials (i.e., a flower, an ice-cream) but they were not depicted, instead the object's locations were marked with a red "x" that measured 15x15 pixels. Here, to adapt the task to the young population, child cartoon characters were created to depict the protagonist and the agent, and 15x15px cartoon objects were also created. The stimuli combined female and male characters and was not gender-matched to the participant. The size and position of the sandbox in the slides followed the exact measurements of Samuel et al. (2018a). The scenarios were created with Adobe Photoshop and converted into landscape PNG files of 756 (width) x 567 (depth) pixels. In each scenario, the sandbox was positioned in the middle on the horizontal axis, and it measured 584x36 pixels.

The initial location of the object as well as the distances between locations in each trial were determined following Samuel et al.'s (2018a) measurements. They used two types of distances between locations: a short-distance move, and a long-distance move. The long-distance move was applied to all the trials because the short-distance move would have been too complicated for children. In each trial the object was moved 45.3% of the sandbox either to the left or to the right of location A (see the distance measurements and direction of move in Table 13). The direction of the move was counterbalanced and the answer to the correct location differed between trials.

Four different within-participant trials were created: the standard FB trial (Figure 18), the altercentric trial (Figure 19), the TB trial (Figure 20) and the prank trial (Figure 21).These four trial types were created in four different between-participant conditions: in the ingroup-ingroup condition both protagonist and agent belonged to the participant's team (Figure 18); in the ingroup-outgroup condition, the protagonist belonged to the participant's team and the agent belonged to the other team (Figure 19); in the outgroup-ingroup condition the protagonist belonged to the other team and the agent belonged to the participant's team (Figure 20) ; and in the outgroup-outgroup condition, both protagonist and agent belonged

to the other team (Figure 21). Sixteen different trials and a total of 128 different images (eight still pictures per trial) were created.

Table 13. Sandbox task exact locations and distances per trial type.

Trial Type	Location A (cm)	Location B (cm)	Distance %	Location A (px)	Location B (px)	Direction
Standard FB	30.1	75.4	45.3	261.784	526.336	Right
Altercentric	80.2	34.9	45.3	554.368	289.816	Left
TB	33.4	78.7	45.3	281.056	545.608	Right
Prank	70.2	24.9	45.3	495.968	231.416	Left

Note: the locations in cm were extracted from Samuel et al. (2018a), the locations in pixels are the exact locations converted from cm to px that were used for this study. Highlighted in brown are the correct trial locations.

Bias calculation for the Sandbox task

Samuel et al.'s (2018a) bias calculation was followed. In each trial, the pixel on which the participant clicked (the x-coordinate) was converted into a measure of the participant's relative bias. In standard FB trials the bias was obtained from taking the difference between the pixel the participant clicked on and the pixel of the correct location (Location A) and dividing the difference by the length of the sandbox. A positive bias meant the participant's location was closer in the direction of Location B (incorrect). In altercentric trials, the bias was obtained from taking the difference between the pixel the participant clicked on and the pixel of the correct location (Location B) and dividing the difference by the length of the sandbox. A positive bias meant the participant's location was closer in the direction of location A (incorrect). In true belief trials, the bias was obtained from taking the difference between the pixel of the correct location (Location B) and the pixel the participant clicked on and dividing the difference by the length of the sandbox. A positive bias meant the participant's location was closer in the direction of Location A(incorrect). Lastly, in prank trials, the bias was obtained from taking the difference between the pixel of the correct location (Location A) and the pixel the participant clicked on and dividing the difference by the length of the sandbox. A positive bias meant that the participant's location was closer in the direction of Location B(incorrect).

Figure 18. Recreation of the FB trial in the ingroup-ingroup condition.

Figure 19. Recreation of the altercentric trial in the ingroup-outgroup condition.

"Here is one kid on the other team and one kid on your team.
Molly and Olivia are at the playground."

"Molly takes her dice…"
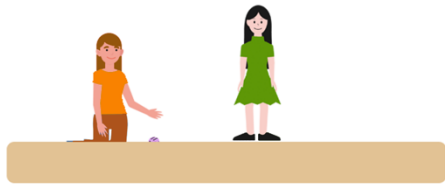
"…she hides it in the sand here…"
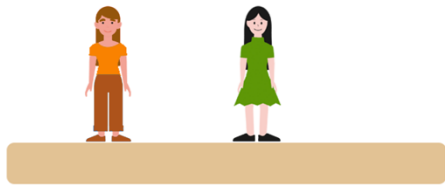
"…and stays."

"While Molly is watching, Olivia finds the dice…"

"…she hides it here…"

"…and she leaves."

"Where does Molly think her dice is?"

Figure 20. Recreation of the TB trial in the outgroup-ingroup condition.

"Here is a kid on the other team. Sam is at the playground."

"Sam takes his sticker…"

"…he hides it in the grass here…"

"…and he goes inside."

"While Sam is inside, the wind starts blowing…"

"…and moves the sticker here."

"Let's play a funny trick on Sam! Let's hide a rubber bug where he thinks the sticker is!"

"Where should we hide the rubber bug?"

Figure 21. Recreation of the prank trial in the outgroup-outgroup condition.

Social Activity FB Task

In this task by Nguyen and Frye (1999) the participant sees the protagonist and the agent conducting an activity together (Activity A). The protagonist leaves and while they are gone, the agent switches to another activity (Activity B). In a forced-choice question, participants are asked about what the protagonist thinks the agent is doing in the other room: Activity A or Activity B. In this adaptation of this task, four different within-participant trials were created: the Individual Neutral Activity trial in which Activity B was a neutral activity (Figure 22); the Individual Mischievous Activity trial in which Activity B was a mischievous activity (Figure 23); the Group Neutral Activity trial in which there was a protagonist and three other agents and Activity B was a neutral activity (Figure 24) and the Group Mischievous Activity trial in which there was a protagonist and three other agents and Activity B was a mischievous activity (Figure 25). As in the Sandbox task, four between-participant conditions of each trial were created: in the ingroup-ingroup condition (Figure 22); the ingroup-outgroup condition (Figure 23); the outgroup-ingroup condition (Figure 24) and the outgroup-outgroup condition (Figure 25). Sixteen different trials and a total of 80 di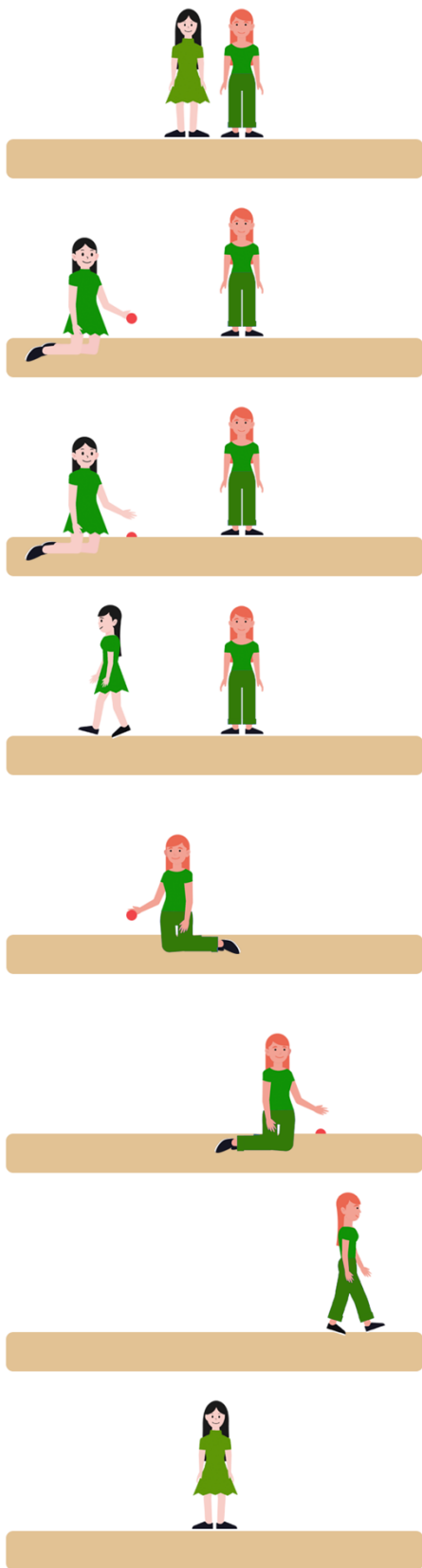fferent images (five still pictures per trial) were created. All child cartoon characters and scenarios were digitally created. The order of presenting the activities in the test question was counterbalanced across trials. *Coding.* Children were given 1 point if they pass the task and 0 if they failed.

Explicit Preferences Task

The stimuli of the Explicit Preferences task (Qian et al., 2016; Setoh et al., 2019) was the same that was used in Studies 2 and 4. Participants had to choose between minimal group ingroup team members and outgroup team members in five different scenarios (5 trials). The position of the characters was counterbalanced across trials.

Coding. They were given one point per trial if they chose an ingroup member and the scores were computed as the proportion of trials were the participant chose an ingroup member.

"Here are two kids on your team. Polly and Liv are reading a book together."

"Now Polly leaves to get a glass of milk."

"Liv doesn't want to read anymore, she starts playing with a teddy bear."

"Now Polly is drinking milk in the kitchen."

"What does Polly think Liv is doing in the other room? [Reading a book] or [playing with the teddy bear]?"

Figure 22. Recreation of the Individual Neutral Activity trial in the ingroup-ingroup condition.
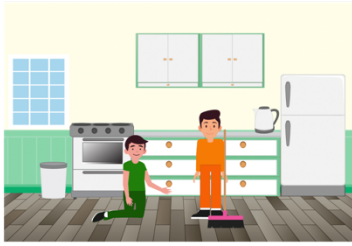
"Here is one kid on your team and one kid on the other team. Tom and Mike are cleaning the kitchen together."

"Now Tom leaves to blow his nose."

"Mike doesn't want to clean anymore, he starts baking and making a mess."
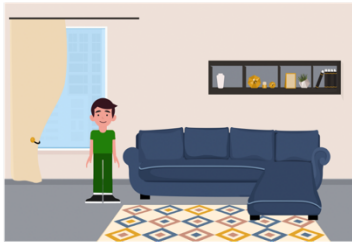
"Now Tom is blowing his nose in the living room."

"What does Tom think Mike is doing in the other room? [Cleaning] or [baking and making a mess]?"

Figure 23. Recreation of the Individual Mischievous Activity trial in the ingroup-outgroup condition.

"Here is a kid on the other team and some kids on your team. They are playing basketball."

"Now Jake leaves to get a towel."

"The kids on your team don't want to play basketball anymore, they start playing soccer."
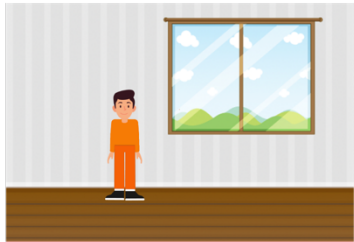
"Now Jake is drying his sweat with the towel at the locker room."

"What does Jake think the kids on your team are doing in the other room? [Playing basketball] or [playing soccer]?"

Figure 24. Recreation of the Group Neutral Activity trial in the outgroup-ingroup condition.

"Here are some kids on the other team. They are coloring books."

"Now Mary leaves to wash her hands."

"The kids on Mary's team don't want to color anymore, they start ripping the pages and making paper planes."

"Now Mary is washing her hands in the bathroom."

"What does Mary think the kids on her team are doing in the other room? [Coloring] or [making paper planes]?"

Figure 25. Recreation the Group Mischievous Activity trial in the outgroup-outgroup condition.

Procedure

Participants were tested individually by an experimenter during a moderated online session via Zoom. The experimenter first interacted with the parents of the participant to obtain audio consent before the testing session. Parents were asked to be present throughout the study and only intercede if the child needed their technical assistance. Experimenters communicated with participants through audio and video streaming and presented the stimuli via a shared screen. The Sandbox task required that the participant clicked on a location on the screen on their own.

Children were first placed in a team through the MGP assignment. Then, participants were randomly assigned to one of the four conditions. They completed eight trials, four Sandbox trials and four Social Activity FB trials. The two tasks were presented in an alternating order, and, following Samuel et al.'s (2018a) procedure, the same order was maintained throughout the conditions: (1) Standard FB trial, (2) Individual Neutral Activity trial, (3) Altercentric trial, (4) Individual Mischievous Activity trial, (5) True Belief trial, (6) Group Neutral Activity, (7) Prank trial and (8) Group Mischievous Activity trial. After the eight trials, children's ingroup bias was tested with the Explicit Preferences Task.

The three tasks were programmed in Qualtrics. For the Sandbox task the Qualtrics' Heatmap question was used. This type of question presents a picture, it allows the participant to click anywhere on the image and saves the exact coordinates of the pixel on which the participant has clicked. In the Sandbox trials, the experimenter enabled Zoom's remote-control feature so the participant could answer the test question by clicking on their own screen with the mouse or trackpad. To ensure that children were technically capable of clicking on a spot of their choosing on the screen, a calibration slide was included right after the MGP assignment and before the start of the trials.

The testing sessions lasted around 30 minutes, depending on how much time the child needed to master the remote-control clicking feature. Participants were tested in their native language, English.

5.1.3. Results

Sandbox task

Following previous bias analysis of the sandbox task (Ni et al., 2022; Samuel et al., 2018b), one-sample *t*-tests were first conducted to measure whether the mean scores in each individual trial were different than zero. As indicated above, a positive result would indicate children's bias toward the incorrect location. Table 14 depicts all *t*-tests by type of trial and condition. In standard FB trials, children's bias score was significantly greater than zero in the two conditions where protagonist and agent shared group-membership: ingroup-ingroup and outgroup-outgroup. Children in all conditions presented a significant bias score towards the incorrect location in the TB trials. Figure 26 presents a graphical representation of the data by type of trial and condition.

Table 14. One-sample t-tests against zero by trial type and condition (*n*=16 per condition).

Condition	Trial Type	Estimates	CI	*p*
Ingroup-Ingroup	**Standard FB**	**0.25**	**0.10 – 0.39**	**0.002**
	Altercentric	0.08	-0.04 – 0.21	0.177
	TB	**0.19**	**0.05 – 0.33**	**0.013**
	Prank	-0.00	-0.09 – 0.09	0.931
Ingroup-Outgroup	Standard FB	0.11	-0.02 – 0.23	0.083
	Altercentric	0.09	-0.05 – 0.22	0.182
	TB	**0.16**	**0.01 – 0.31**	**0.034**
	Prank	0.00	-0.11 – 0.12	0.963
Outgroup-Ingroup	Standard FB	0.10	-0.04 – 0.24	0.157
	Altercentric	0.08	-0.06 – 0.22	0.259
	TB	**0.19**	**0.04 – 0.33**	**0.016**
	Prank	0.07	-0.08 – 0.22	0.355
Outgroup-Outgroup	**Standard FB**	**0.23**	**0.08 – 0.39**	**0.006**
	Altercentric	-0.03	-0.11 – 0.04	0.321
	TB	**0.14**	**0.02 – 0.27**	**0.025**
	Prank	-0.04	-0.12 – 0.04	0.280
All conditions together	**Standard FB**	**0.17**	**0.10 – 0.24**	**<0.001**
	Altercentric	0.05	-0.00 – 0.11	0.065
	TB	**0.17**	**0.11– 0.23**	**<0.001**
	Prank	0.01	-0.05 – 0.06	0.810

Note: Highlighted in bold are the trials in which participants were biased towards the incorrect location.
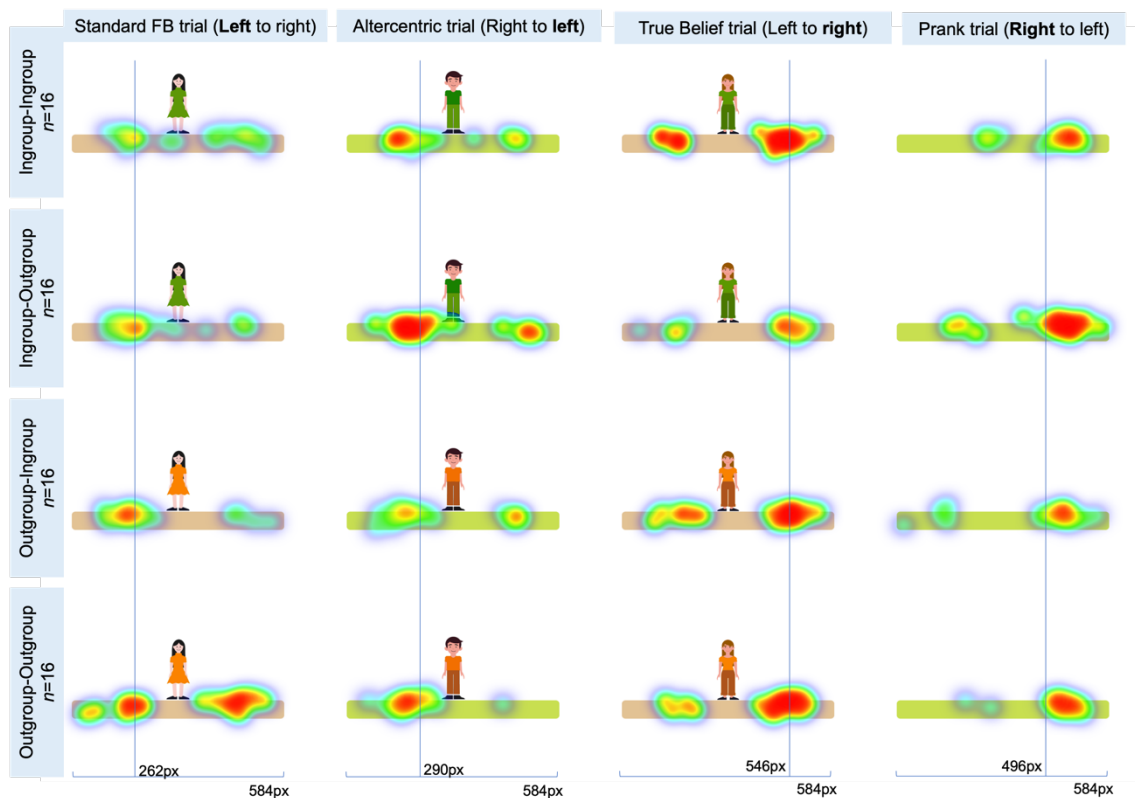
Figure 26. Qualtrics Heat Map Plot by trial type and condition. The pixel coordinates of the participants' responses are overlaid across the images with colors that indicate the frequency of respondent's selections. A redder area means that more participants clicked on that pixel. The blue line indicates the precise correct location of the object. Images are presented with Qualtrics' default values for radius and blur visualization.

To assess whether biases differed between conditions a mixed measures ANOVA with trial type as a within-subject factor and group membership as a between-subject factor was carried out. This analysis yielded a main effect of Trial Type, $F(3,186) = 6.91$, $p < .001$, $\eta_p^2 = .03$, but no condition or interaction effect. Tukey HSD post-hoc comparisons of the four conditions together showed that the bias in the standard FB trial ($M=0.12$, $SD=0.26$) was significantly larger than the bias in the altercentric trial ($M=0.08$, $SD=0.24$, $p=.03$) and the prank trial ($M=0.02$, $SD=0.22$, $p<.001$). It was also revealed that the bias in the true belief trial ($M=0.18$, $SD=0.25$) was significantly larger than the bias in the altercentric trial ($p=.031$) and the prank trial ($p<.001$).

Sandbox task timing

During the testing sessions the Qualtrics time keeping feature in the sandbox trials test slides was activated. See Table 15 for a summary of the Page submit time in each trial type by condition.

Table 15. Page submit time (in seconds) per trial and condition.

Condition	Trial Type	Mean	SD
Ingroup-Ingroup	Standard FB	21.99	16.80
	Altercentric	15.74	11.16
	TB	11.84	7.61
	Prank	11.01	6.79
Ingroup-Outgroup	Standard FB	16.24	7.12
	Altercentric	14.05	14.37
	TB	12.22	13.50
	Prank	7.27	3.06
Outgroup-Ingroup	Standard FB	17.81	9.80
	Altercentric	14.20	5.12
	TB	12.79	6.71
	Prank	10.86	6.10
Outgroup-Outgroup	Standard FB	18.75	11.92
	Altercentric	15.98	12.22
	TB	10.70	11.24
	Prank	10.53	6.51
All conditions together	Standard FB	18.70	11.85
	Altercentric	15.10	11.01
	TB	11.88	9.93
	Prank	9.92	5.87

As an exploratory analysis, a mixed measures ANOVA with trial type as a within-subject factor and group membership as a between-subject factor was conducted. The analysis revealed a significant main effect of trial, $F(3,186) = 17.77$, $p < .001$, $\eta_p^2 = .00$. Tukey HSD post-hoc comparisons showed that the page submit time in the standard FB trial was significantly longer than in the TB trial ($p<.001$) and the prank trial ($p<.001$). It was also revealed that the page submit time in the altercentric trial was significantly longer than in the prank trial ($p=.018$).

Social Activity FB task

One-sample *t*-tests were first conducted to measure whether the mean scores in each trial differed from chance. Table 16 depicts all *t*-tests by type of trial and condition. There were only three trials across conditions in which children's performance was at chance: the Individual Neutral Activity trial in the ingroup-ingroup condition, and the Individual Neutral Activity and the Group Neutral Activity in the ingroup-outgroup condition.

A mixed measures ANOVA with trial type as a within-subject variable and group membership as a between-subject variable yielded no significant effects.

Table 16. One-sample t-tests against chance by trial type and condition (*n*=16 per condition).

Condition	Trial Type	Estimates	CI	*p*
Ingroup-Ingroup	**Individual Neutral Activity**	**0.62**	**0.36-0.89**	**0.33**
	Individual Mischievous Activity	0.81	0.60-1.03	<.01
	Group Neutral Activity	0.81	0.60-1.03	<.01
	Group Mischievous Activity	0.81	0.60-1.03	<.01
Ingroup-Outgroup	**Individual Neutral Activity**	**0.69**	**0.43-0.94**	**0.138**
	Individual Mischievous Activity	0.81	0.60-1.03	<.01
	Group Neutral Activity	**0.69**	**0.43-0.94**	**0.138**
	Group Mischievous Activity	0.81	0.60-1.03	<.01
Outgroup-Ingroup	Individual Neutral Activity	0.88	0.69-1.06	<.001
	Individual Mischievous Activity	0.75	0.51-0.99	0.041
	Group Neutral Activity	0.75	0.51-0.99	0.041
	Group Mischievous Activity	0.88	0.69-1.06	<.001
Outgroup-Outgroup	Individual Neutral Activity	0.75	0.51-0.99	0.041
	Individual Mischievous Activity	0.75	0.51-0.99	0.041
	Group Neutral Activity	0.81	0.60-1.03	<.01
	Group Mischievous Activity	0.81	0.60-1.03	<.01
All conditions together	Individual Neutral Activity	0.73	0.62-0.85	<.001
	Individual Mischievous Activity	0.78	0.68-0.89	<.001
	Group Neutral Activity	0.77	0.66-0.87	<.001
	Group Mischievous Activity	0.83	0.73-0.92	<.001

Note: Highlighted in bold are the trials in which participants did not perform different than chance.

Explicit preferences task

All children showed an explicit bias favoring their ingroup team members, $t(63) = 49.74$, $p < .001$, $d=6.22$, and the bias did not differ between group membership conditions, $F(1,62) = 0.75$, $p = .39$, $\eta_p^2 = .01$.

5.1.4. Discussion

The goal of this study was to test the pragmatic influence on performance of changing the identity of the two agents in a continuous FB task. Participants were assigned to a team through a MPG, and four different conditions were created combining the ingroup-outgroup identities of the protagonist-agent dyad. Children first completed the standard FB trial child-friendly adaptation of the Sandbox task, and then they were presented with three additional Sandbox trials that were alternated with four different trials of the Social Activity FB task.

As in Samuel et al.'s (2018a) study, children's bias scores were calculated by taking the difference in pixels between the correct location and the location the participant clicked on and dividing it by the length of the sandbox. A positive score that significantly differed from zero would show that children were biased towards the wrong location through the manipulation in the study. The bias scores were first analzyed in each individual trial.

The data revealed that, children's bias scores in the standard FB trial were only significantly greater than zero in the conditions in which the protagonist and the agent shared group membership. These results are in line with the hypothesis that children would expect a protagonist that belonged to the same group as the agent to be more likely to share the agent's knowledge; it was predicted that the protagonist's expected knowledge would function as a pragmatic trigger to bias children's responses. The data also revealed that children were biased towards the wrong location in all TB trials, and they were not biased in any of the altercentric trials or prank trials.

The analysis detected a significant effect of trial type on performance, but it did not show a condition effect or an interaction effect between condition and trial type. These results replicate previous findings from Samuel et al. (2018b) and Ni et al. (2022) where no difference between trial conditions was found using an online adaptation of the Sandbox

task. However, in these studies, children exhibited a positive bias score significantly greater than zero in all the trial types[21], which was not the case in this study.

It had been predicted that children's positive bias scores across conditions would be the largest in TB trials, followed by FB trials. It was also predicted that prank trials would not exhibit a positive bias score. There was no clear prediction for children's performance in the altercentric trials in relation to the other three trials. The study revealed that children across conditions exhibited a positive bias in the TB tasks as well as in the FB tasks, and the bias was not significantly larger in the TB condition than in the FB condition. Performance in these two trial types did significantly differ from performance in the altercentric trials and the prank trials; children were not influenced by the presence of the protagonist when they were asked about the current location of the object and framing the FB trial as a prank was proved to decrease children's bias scores.

The exploratory analysis on timing revealed that children were faster to respond in the TB task and in the prank trial than in the FB and altercentric trials. This difference could very well reflect a practice effect since it was the last two trials the ones that required the less amount of time to respond. It is interesting to note that the fact that they required the less amount of time only meant proficiency in response in the case of the prank trial. Descriptively, children's response time was the slowest in the FB trials of the two shared-group membership conditions.

Finally, regarding the Social Activity FB task, children's performance in this task was significantly above-chance in most of the trials and the results did not reveal an effect of trial type, condition, or an interaction between the two in performance of the Social Activity FB task. I take these findings to point to the inadequacy of this task for this age range. The prediction was that shared-group membership between the protagonist and the agent, or the

---

[21] Samuel et al. (2018b) tested adults' egocentric bias in three different test trial types: a memory trial in which they reasoned about another's memory ("where does the protagonist *remember* [the object] to be?"), a FB trial in which they reasoned about another's belief ("where does the protagonist *believe* [the object]to be?" and an action prediction trial where they had to point to where the protagonist would *look for* the object. They found a positive bias score significantly greater than zero in the three experimental conditions as well as in the control condition (where they reasoned from their own memory). Ni et al. (2020) tested children's automatic perspective-taking with three between-group conditions (joint attention, presence of an experimenter, absence of the experimenter) and found a positive bias that was significantly greater than zero in all conditions.

protagonist and the participant would increase the pragmatic difficulty of the task. The data partially supported this prediction. Children's performance in the Individual Neutral Activity trial was at chance in the ingroup-ingroup condition providing confirmation that the shared-group membership between protagonist-agent-participant made the task pragmatically harder. The other two at-chance performances were from both neutral trials in the ingroup-outgroup condition providing evidence that the shared group membership between participant and protagonist further influenced the participant's FB attribution as had been seen in the previous studies with the unexpected-contents FB task.

Chapter 6.
Developmental trajectory of the pragmatic manipulation of the protagonist's expected ignorance

6.1. Study 6

6.1.1. Introduction

Due to the dichotomic forced-choice response of the standard FB task, its use is limited to a certain age range. To assess ToM skills with children that have already mastered the standard FB task, researchers have mostly resorted to second-order false belief tasks that are presented either independently (e.g., The Ice-Cream Story - Perner & Wimmer, 1985; the Birthday Puppy test – Sullivan et al. 1994) or as part of a battery of tasks like the Strange Stories task (White et al., 2009) that also includes double bluffs, white lies and jokes. In a second-order FB task participants are tested on their capacity to attribute second-order (or embedded) mental states: "she thinks that he thinks" (Sullivan et al., 1994). The ability to embed mental states inside other mental states (e.g., hold beliefs about beliefs about beliefs) is called recursive mindreading and it can potentially be tested to a high a level of recursive degrees (O'Grady et al., 2015). The cognitive demand that these tasks require differs from the first-order FB tasks because the participant's perspective is not directly pitted against the protagonists' perspective. As reviewed in the introduction, past research in Intergroup ToM revealed that 9-to-13-year-olds' performance was better when mentalizing about an ingroup target in situations in which there was no conflict between the participant's and the protagonist's perspectives (Gönültas et al., 2020).

The aim of this study was to explore whether the manipulation of the agent's identity in the Sandbox task could also influence performance of older children that have been proficient at the standard change-of-location task for a few years already. Since the Sandbox task has been proposed as a measure that can test mindreading abilities in a broader age-range than any standard FB task (Beeger et al., 2012; Bernstein et al., 2011; Bernstein et al., 2017; Mahy et al., 2017; Sommerville et al., 2013) I tested the pragmatic manipulation's influence in 8-to-9-year-olds' performance using the Sandbox task.

6.1.2. Method

Participants

Seventy-four 8-to-9-year-old children participated in the study[22]. They were recruited from the SCD Lab at Yale's database, via social media ads, the *Children Helping Science* platform, and at recreational parks in the New Haven area. Eleven children were excluded because they were not able to complete the assignment. The final sample consisted in 64 children (35 female, M_age= 107 months; SD_age: 7 months) that were randomly assigned into four conditions (16 children per condition). Children belonged to middle-class families. Racial data was not collected. The study was approved by the institutional review board of the university and a caregiver's informed consent was obtained for each participant.

Measures and procedure

Children were first assigned to the green team through Study 5's MGP assignment. They were randomly assigned to one of four conditions (ingroup-ingroup, ingroup-outgroup, outgroup-ingroup, outgroup-outgroup) and completed the three measures from Study 5 (Sandbox task, Social Activity FB task, Explicit Preferences Task). Most participants were tested by an experimenter during a moderated online session following the exact procedure as in Study 5. Fifteen participants were tested in-person at the New Haven Green (New Haven, CT, US). These participants completed the studies on an iPad device under the supervision of the experimenter.

6.1.3. Results

Sandbox task

Preliminary analysis yielded no difference between the participants that were tested in person and those that were tested via Zoom, $F(1,248) = 0.35$, $p = .55$, $\eta_p^2 = .00$, so testing location was not included in the analysis. Following Study's 5 analysis plan one-sample t-tests were conducted to measure whether the mean scores in each trial were different than zero. A positive result would indicate children's bias toward the incorrect location. Table 17 shows all t-tests by type of trial and condition. Children were biased towards the incorrect location

---

[22] Our recruiting goal was to have 16 participants per condition. We stopped the testing sessions once we got valid data from 16 participants.

in TB trials in all conditions except the ingroup-ingroup condition. Figure 27 presents a graphical representation of the data by type of trial and condition.

Table 17. One-sample t-tests against zero by trial type and condition (*n*=16 per condition).

Condition	Trial Type	Estimates	CI	*p*
Ingroup-Ingroup	Standard FB	0.05	-0.07-0.17	0.407
	Altercentric	0.10	-0.03-0.23	0.118
	TB	0.12	-0.01-0.24	0.062
	Prank	-0.06	-0.13-0.02	0.132
Ingroup-Outgroup	Standard FB	0.11	-0.04-0.27	0.140
	Altercentric	0.11	-0.03-0.25	0.105
	TB	**0.20**	**0.05-0.35**	**0.013**
	Prank	0.13	-0.03-0.30	0.106
Outgroup-Ingroup	Standard FB	0.13	-0.00-0.26	0.055
	Altercentric	0.11	-0.03-0.27	0.114
	TB	**0.13**	**0.01-0.25**	**0.036**
	Prank	0.05	-0.07-0.18	0.383
Outgroup-Outgroup	Standard FB	-0.02	-0.10-0.07	0.706
	Altercentric	0.10	-0.04-0.23	0.151
	TB	**0.32**	**0.18-0.47**	**<0.001**
	Prank	0.00	-0.11-0.11	0.986
All conditions together	**Standard FB**	**0.07**	**0.01-0.13**	**0.024**
	Altercentric	**0.12**	**0.05-0.18**	**<0.001**
	TB	**0.19**	**0.13-0.26**	**<0.001**
	Prank	0.03	-0.03-0.09	0.271

Note: Highlighted in bold are the trials in which participants were biased towards the incorrect location.
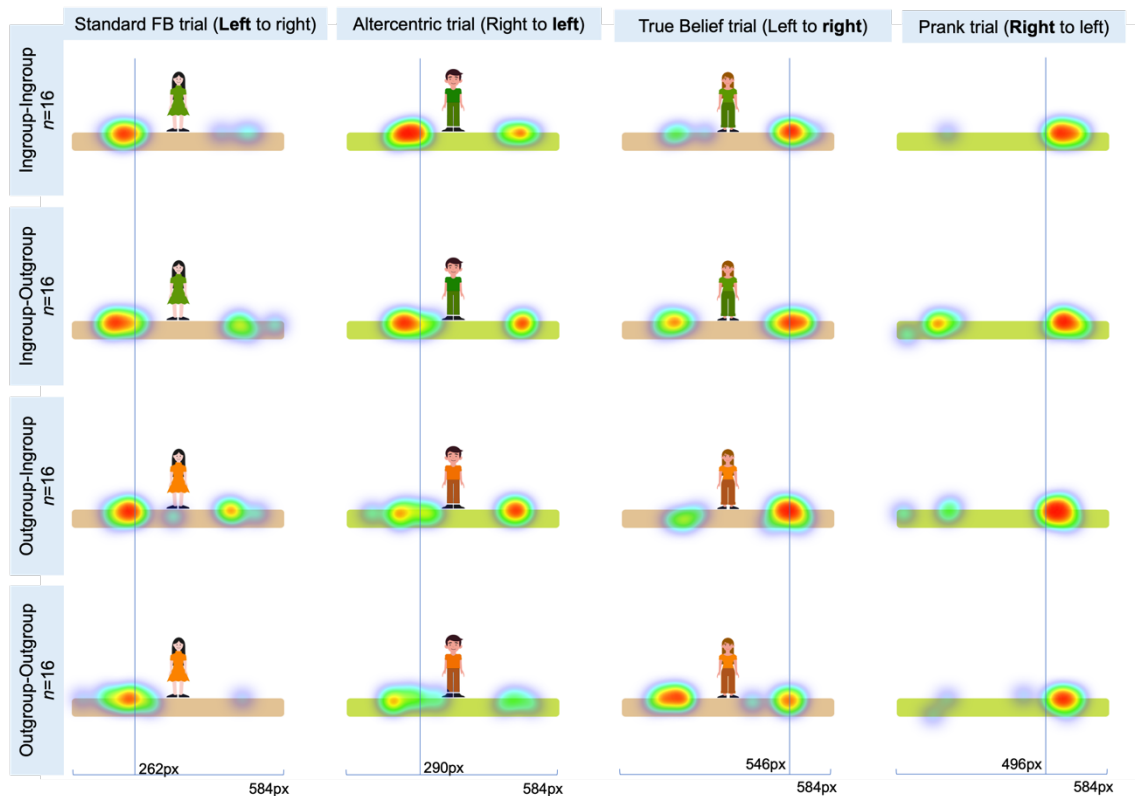
Figure 27. Qualtrics Heat Map Plot by trial type and condition (8-9-year-olds).

To assess whether biases differed between conditions a mixed measures ANOVA with trial type as a within-subject factor and group membership as a between-subject factor was carried out. This analysis yielded a main effect of Trial Type, $F(3,186) = 4.81$, $p = .003$, $\eta_p^2 = .01$, but no condition or interaction effect. Tukey HSD post-hoc comparisons of the four conditions together showed that the bias in the TB trial ($M=0.19$, $SD=0.26$) was larger than the bias in the standard FB trial ($M=0.06$, $SD=0.23$, $p=.027$) and the prank trial ($M=0.03$, $SD=0.23$, $p=.002$).

Social Activity FB task

Participants performed above chance in all trials across conditions. A mixed measures ANOVA with trial type as a within-subject variable and group membership as a between-subject variable did not yield significant effects.

Explicit preferences task

All children showed an explicit bias favoring their ingroup team members, $t(63)= 10.64$, $p < .001$, $d=1.33$, and the bias did not differ between group membership conditions, $F(1,62) = 0.01$, $p = .91$, $\eta_p^2 = .00$.

6.1.4. Discussion

In the present study I wanted to explore the developmental trajectory of the influence in performance of the agents' group membership using a task that maintained a conflicting participant-protagonist perspective with a population that was already proficient in standard FB tasks (8-to-9-year-olds).

Separately evaluating children's bias scores against chance in each trial by condition showed that the agents' group membership did not influence performance in the FB trial. Yet merging the four conditions together showed that children's bias scores were significantly greater than zero in the FB trials. The two agents' group membership manipulation was not strong enough to yield different results in 8-to-9-year-olds' between-group performance in FB trials, yet the Sandbox task was sensitive enough to detect a bias towards the wrong location in FB trials.

The paradigm not only revealed a positive bias in the FB trial, but, as in Study 5, it also revealed a positive bias in the TB trial. Rakoczy and Oktay-Gür (2020) reported that children start to pass both tasks between the ages of 8 and 10 but performance is still strongly negatively correlated until age 10. The TB trial's positive bias in the current study was significantly larger than the bias in the FB trial, indicating that the children in this sample were still in the sensitive age-range in which performance in TB-FB is negatively correlated, even more so than the 5-to-6-year-olds. Expectedly, the bias in TB trials was also significantly different than the bias in prank trials that, in line with the previous study, yielded the most accurate performance.

Surprisingly, the bias score for the TB trials did not differ from the bias score in altercentric trials. Children's overall bias scores in altercentric trials was significantly greater than zero, pointing to an altercentric interference in children's performance - which had not been predicted. One interpretation could be that children's perspective-tracking process was indeed influenced by the presence of the protagonist. This interpretation is not entirely

plausible as the younger children in Study 5 had not shown any altercentric interference. Another possibility could be that the seeming easiness of the test question, a reality-based question "Where is the car now?", was pragmatically confusing for participants in the same way that the TB task is pragmatically hard for children in this age range. Resorting to the predictions of Rakoczy and Oktay-Gür's (2020) Pragmatic Performance Limitation Account, children in this sample might have not had the "higher-order, open-ended pragmatic flexibility of adults" (p.3) that would be needed to make sense of this specific speech act (the experimenter's reality question).

As expected, children's performance in the Social Activity FB task was well above chance and all participants showed an explicit preference for their ingroup team members.

6.2. Study 7

6.2.1. Introduction

The goal of Study 7 was to further explore the developmental trajectory of the influence in performance of the group membership manipulation using a task that pitted the protagonist's perspective against the participant's, with a population that was even more proficient in FB attribution than the one in Study 6, adults.

The field of adult intergroup perspective-taking (i.e., a broader understanding of ToM) is considerably more established than the developmental counterpart. Though the breadth of research is more substantial, the influence of the mentalizing target seems to follow a similar pattern to children's studies. When the measure used requires the participant to project a mental state in a situation that does not involve a conflicting perspective, there seems to be an ingroup advantage in mentalization. Perez-Zapata et al. (2016) and Ekerim-Akbulut et al. (2019) used the Strange Stories Task (White et al., 2009) and found that participants were more accurate in their mental-state understanding with an ingroup target. This advantage, as it happens with children, seems to be correlated with a better perceptual understanding of the ingroup's emotions (Adams et al. [2010] reported that adults showed an ingroup advantage in Baron-Cohen et al.'s 2001 "Reading the Mind in the Eyes Test" performance) and mind perception (Hackel et al. [2014] found that adults' had outgroup targets required more humanness to be perceived as having minds as opposed to ingroup targets).

In paradigms where the perspectives are against one another, researchers have reported an outgroup advantage. Savitsky et al. (2011) described a "close-communication bias" that worsened adults' performance when they completed an adaptation of Keysar et al.'s (2000) Director Task with a friend versus a stranger. Yet, the most thorough evidence for an outgroup advantage was provided by Todd et al. (2011). The author's hypothesis was that a difference mind-set ("a cognitive orientation wherein distinctive self-referential information is activated and used as a comparison standard to draw inferences about other people," p.135) would enhance adults' performance in perceptual and conceptual forms of perspective-taking. In the first three studies, the difference-mindset was artificially generated with a priming procedure. They found an advantage for the experimental condition (in which a difference-mindset had been primed) in a spatial perspective-taking task (Tversky & Hard,

2009), a message-interpretation task (Keysar, 1994) and in an adaptation of Birch and Bloom's (2007) false-belief task for adults[23]. In the last two studies the difference-mindset was generated by the "outgroupness" of the target. In one study they tested adults' performance in Birch and Bloom's (2007) change-of-location FB task towards an ingroup target (German) and an outgroup target (Turkish) and found a better performance in the cultural outgroup target condition. In the last study they used a MGP to divide participants into two random conditions and tested adult's performance in a maze direction task (Stephenson & Wicklund, 1984). They also found a performance advantage in this task. Lastly, Simpson and Todd (2017) used the Dot task with two group membership manipulations (university affiliation and MGP) and found that the interference from the participant's perspective was stronger when responding from an ingroup versus an outgroup's perspective.

Evidence for adult bias scores in the Sandbox task has been mixed. Whereas the pen and pencil versions of the task showed evidence for a greater adult bias in a FB trial compared to an own memory trial in which the participant is asked where the protagonist left the object (Bernstein et al., 2011; Bernstein et al., 2017; Mahy et al., 2017; Sommerville et al., 2013), the online version of the task has provided unconclusive results. Samuel et al. (2018a) found that adults were more biased towards the wrong location reasoning from the protagonist's false belief than from their own memory. Samuel et al. (2018b) could not replicate the finding and concluded that the Sandbox task was not "sensitive enough to draw out consistent effects related to mental state reasoning in young adults" (p.1).

In the current study the aim was to explore the influence of the shared group membership pragmatic manipulation with an adult population and determine whether the Sandbox task would be a sensitive enough measure to detect differences between groups and trial types in adults' performance.

---

[23] In this change-of-location task there were four colored containers (blue, red, purple, green). The protagonist left the object in the blue container and left. Then, participants in the informed condition are told that while the protagonist is away the agent moves the object to the [red] container; participants in the ambiguous condition are told that the agent moves it to [another] container. The agent then rearranges the containers in the room so that the red container is where the blue container originally was. Participants are asked "What are the chances that the protagonist will look for the object in each of the above containers?" and need to write their answer in % beneath each container. Participants primed with the difference-mindset were less biased than control participants in the informed condition (when they knew the location).

6.2.2. Method

Participants

A hundred and twenty-eight American adults (M_age= 35 years; SD_age: 8.5 years) participated in the study. They were recruited through Turk Prime and tested using Qualtrics. They were randomly assigned into four conditions (32 participants per condition). The study was approved by the institutional review board of the university and an informed consent was obtained for each participant.

Measures and procedure

Participants were assigned to the green team following the same procedure of the two previous studies. They were randomly assigned to one of four conditions (ingroup-ingroup, ingroup-outgroup, outgroup-ingroup, outgroup-outgroup) and completed an adaptation of Study's 5 Sandbox task, the Social Activity FB task and the Explicit Preference Task.

To adapt the child-friendly version of the Sandbox task to an adult population the example of Samuel et al. (2018a) was followed: a word search puzzle was added right before the test slide to create a distraction to deter participants from using a perceptual strategy to answer the question. Seven by seven word-search puzzles were created (Figure 28) using an online puzzle maker software (www.puzzlemaker.com). These puzzles were displayed during 15 seconds between the last slide of the scenario and the test question slide and participants were encouraged to find as many words as possible.

```
S  W  D  Q  N  R  L
C  V  I  R  E  M  C
H  Y  H  H  A  H  L
A  C  T  O  R  M  L
P  R  O  T  O  M  A
V  Z  Z  E  D  V  F
E  Q  Y  L  H  W  A
```

Figure 28. Example of word search puzzle distractor for the adult version of the Sandbox task.

6.2.3. Results

Sandbox task

Following Studies 5 and 6' analysis plan one-sample t-tests were conducted to measure whether the mean scores in each trial were different than zero. A positive result would

indicate children's bias toward the incorrect location. Table 18 shows all *t*-tests by type of trial and condition. Participants were biased towards the incorrect location in TB trials in all conditions except the outgroup-ingroup condition. Moreover, they were biased in altercentric trials in all conditions. Figure 29 presents a graphical representation of the data by type of trial and condition.

Table 18. One-sample t-tests against zero by trial type and condition (*n*=32 per condition).

Condition	Trial Type	Estimates	CI	*p*
Ingroup-Ingroup	Standard FB	0.04	-0.02-0.11	0.205
	Altercentric	**0.10**	**0.00-0.19**	**0.042**
	TB	**0.09**	**0.01-0.17**	**0.022**
	Prank	0.01	-0.06-0.07	0.837
Ingroup-Outgroup	Standard FB	-0.00	-0.06-0.05	0.884
	Altercentric	**0.09**	**-0.02-0.12**	**0.034**
	TB	**0.14**	**0.07-0.21**	**0.001**
	Prank	-0.03	-0.09-0.03	0.293
Outgroup-Ingroup	Standard FB	-0.02	-0.06-0.02	0.330
	Altercentric	**0.15**	**0.05-0.24**	**0.004**
	TB	0.04	-0.01-0.09	0.149
	Prank	-0.00	-0.08-0.08	0.945
Outgroup-Outgroup	Standard FB	-0.02	-0.08-0.04	0.541
	Altercentric	**0.18**	**0.07-0.28**	**0.002**
	TB	**0.07**	**0.00-0.13**	**0.042**
	Prank	0.04	-0.05-0.12	0.360
All conditions together	Standard FB	0.00	-0.03-0.03	0.989
	Altercentric	**0.12**	**0.07-0.16**	**<0.001**
	TB	**0.08**	**0.05-1.12**	**<0.001**
	Prank	0.00	-0.03-0.04	0.890

Note: Highlighted in bold are the trials in which participants were biased towards the incorrect location.
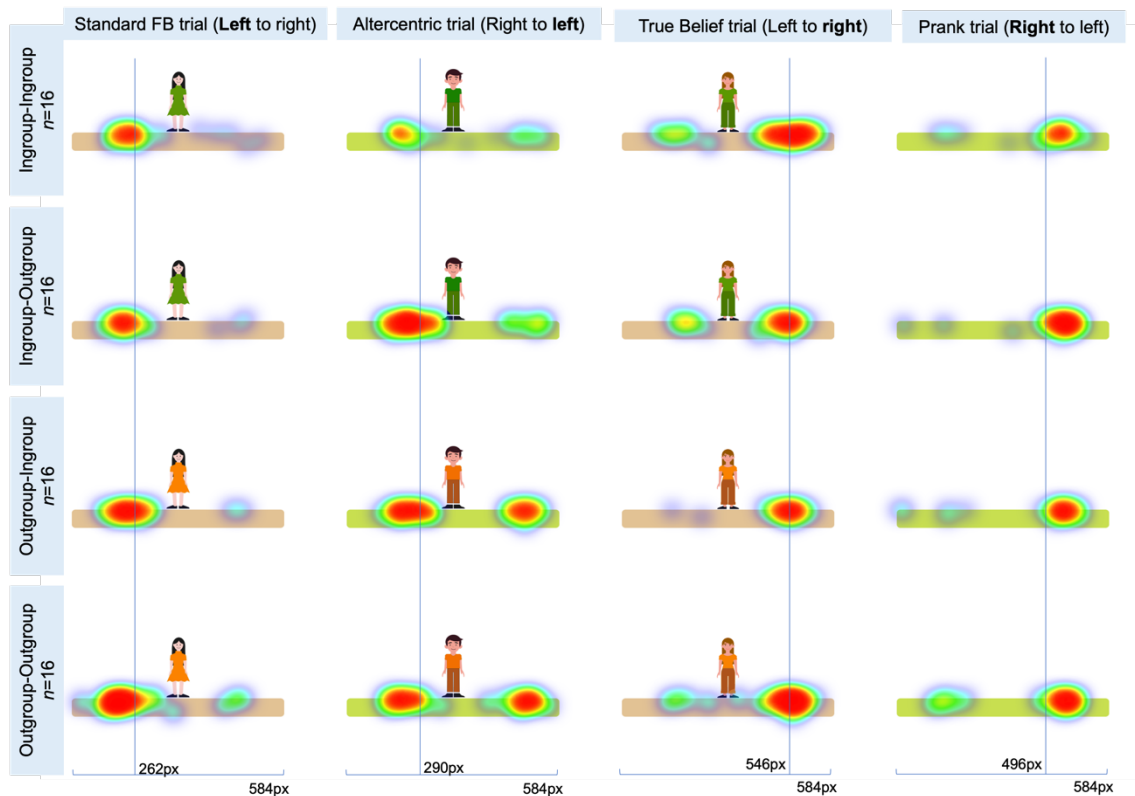
Figure 29. Qualtrics Heat Map Plot by trial type and condition (adult population).

To examine whether bias scores were different depending on the group-membership condition a mixed measures ANOVA with trial type as a within-subject factor and group membership as a between-subject factor was conducted. This analysis yielded a main effect of Trial Type, $F(3,378) = 10.70$, $p < .001$, $\eta_p^2 = .00$, but no condition or interaction effect. Tukey HSD post-hoc comparisons of the four conditions together showed that the bias in the TB trial ($M=0.08$, $SD=0.18$) was larger than the bias in the standard FB trial ($M=0.00$, $SD=0.16$, $p=.007$) and the prank trial ($M=0.02$, $SD=0.20$, $p=.009$). The bias in the altercentric trial ($M=0.11$, $SD=0.26$) was also larger than the standard FB trial ($p<.001$) and the prank trial ($p<.001$).

Social Activity FB task

As in Study 6, participants performed above chance in all trials across conditions. A mixed measures ANOVA with trial type as a within-subject variable and group membership as a between-subject variable yielded non-significant effects.

Explicit preferences task

All participants showed an explicit bias favoring their ingroup team members, $t(127) = 28.14$, $p < .001$, $d=2.49$, and the bias did not differ between group membership conditions, $F(1,126) = 2.25$, $p = .13$, $\eta_p^2 = .02$.

6.2.4. Discussion

Participants did not show a positive bias in the FB trial in any of the conditions. Surprisingly, adults' performance in the TB trial in three of the conditions were significantly biased towards the wrong location. The only condition in which they were not biased was the outgroup-ingroup condition. Though it was not expected that a TB bias would be seen in an adult population, the finding that participants in the outgroup-ingroup condition were the only ones to not show a significant bias was not surprising. According to my account, in this condition, the target's ignorance expectation should be the highest since the protagonist was an outgroup to both the participant and the agent.

The even more surprising finding was to find a positive bias score greater than zero in the altercentric trials of the four conditions. Further, adults were significantly more biased in the altercentric and TB trials than in the FB and prank trials. It was expected the pragmatic difficulty of the trivial TB question to affect children, but it was not expected that it would also influence adults' performance. Moreover, to find an altercentric interference in adults but not in 5-to-6-year-olds was not expected.

The Sandbox task appeared to be not sensitive enough to detect performance differences of the group-manipulation (aside from the small confirmation from TB performance in the outgroup-ingroup condition). Nonetheless, unlike in Samuel et al. (2018b), it revealed to be sensitive enough to detect differences in performance between different trial types. In Samuel et al. (2018b) regardless of the question, participant's performance in all trials were biased towards Location A. There is formatting factor that could explain why adult performance in the trials in the present study was not biased towards Location A by default; the child-friendly adaptation included cartoon characters (instead of only referring to the characters in the text) that depicted the narrative of the story. Perhaps the visual aid helped differentiate between the actual questions in each trial. The presence of a character in the altercentric trials might also have enhanced the potential altercentric interference.

In Study 6 two possibilities for interpreting the positive bias scores in altercentric trials were contemplated. One possibility was that children were indeed influenced by the presence of a target in their answer to a reality question. This interpretation conflicted with Study's 5 lack of evidence of an altercentric bias in younger children. The other possibility was that the pragmatic confusing factors of the altercentric trial (i.e., the triviality of the question) mirrored the pragmatic difficulty that the TB trial poses for children that are already proficient at the FB task. According to this interpretation, children's lack of sophisticated open-ended pragmatic flexibility would be responsible for the positive bias scores in the TB trials as well as the altercentric trials. Interestingly, a population that would be assumed to be competent enough to overcome such pragmatic difficulties still showed a bias towards the wrong location in the TB and the altercentric trials.

Chapter 7.
General discussion

During the early preschool years, the development of the concept of "the self" and "the other" has a significant impact on two socio-cognitive processes (ToM and social categorization) that have largely been studied separately. Developing an understanding that there is an individual "me" and an individual "you" is crucial to infer someone's mental states (desires, beliefs, and intentions), which is an indispensable skill for interpersonal relationships. However, when these early developing categorization processes extrapolate the individual distinction ("me" and "you") to a group distinction ("us" versus "them"), the interpersonal relationships become biased with an indiscriminate preference for those that belong to one's group. While categorizing is necessary for infants to represent the world in a less taxing manner (Dunham & Olson, 2016) it often triggers stereotypical representations of individuals, not attending to who they are but rather which group they belong to. In turn, appreciating individuals through their group-imbued qualities instead of their individual qualities can potentially tax the mental state inferencing process.

The slowly growing field of Developmental Intergroup ToM observes the interaction between mental state inferencing processes and intergroup social cognitive processes during development. One branch of the field is concerned with the social implications that ToM abilities bring to intergroup relationships (Glidden et al., 2021; Mulvey et al.,2016; Mulvey et al., 2021). Another branch in this field examines the cognitive implications that derive from mentalizing about someone who belongs to one's group or someone who does not. Past research in the cognitive branch has shown that in developmental paradigms where there is no conflict between perspectives, there seems to be an advantage when mentalizing about an ingroup target (Gönültaş et al., 2019; McLoughlin & Over, 2017). In contrast, the influence of group membership on performance has been mixed in paradigms where the protagonist's perspective competes with the participant's perspective (i.e., FB paradigms).

Children's performance in FB paradigms has been the subject of a heated debate over the last thirty years. The systematic failure of the standard FB tasks before the 4th birthday has

been interpreted in divergent manners: nativist accounts attribute it to the high processing demands of the task in relation to children's underdeveloped executive and attentional resources before that age (e.g., Baillargeon et al., 2010; Leslie et al., 2004; Leslie et al., 2005); empiricist accounts attribute the failure to a lack of a meta-representational concept of belief (Wellman, 2001; Gopnik & Wellman, 1992) and pragmatic accounts attribute children's poor performance to the specific contextual elements of the task that make it pragmatically difficult for children (Helming et al., 2016; Rubio-Fernández & Geurts, 2013; Siegal & Beattie, 1991; Westra, 2017a).

The Outgroup Advantage Pragmatic Account (OAPA) presented in this dissertation is a pragmatic account that attempts to explain children's difficulty in the FB task through the lens of the cognitive branch of Developmental Intergroup ToM. The focus of the account is to show that performance in standard FB tasks is not solely linked to children's competence in attributing false beliefs but rather it is interwoven with pragmatic factors that make the task cognitively demanding. The OAPA establishes that the social group membership of the recipient of the mentalization (the "false-belief holder") is a pragmatic factor that can influence children's performance in a task that directly pits the protagonist's perspective against that of the participant.

Based on the general learnings from intergroup cognition, the account suggests that an outgroup protagonist will be: *categorized* as an outgroup individual, considered *dissimilar* to oneself, *stereotyped* as an outgroup individual (and thus less valued on warmth and competence), *less preferred,* and *less likely to share* the participant's *knowledge.* The learnings from interpersonal cognitive biases in FB paradigms suggest that children that mentalize about an outgroup protagonist will be less biased by their own *curse of knowledge* or *egocentric distortion* and will *project a stereotype template (low-competence, ignorant character-traits)* instead of recruiting a self-template for mentalization. Lastly, the learnings from the pragmatic accounts suggest that an outgroup protagonist will: make the question *more relevant* and *less academic, legitimize* the *experimenter's interest* in the protagonist's belief, and will be *less of a trigger for a helpfulness interpretation.* In sum, it predicts that an outgroup protagonist reduces the pragmatic demands of the FB task because it presents an easier target to attribute false beliefs to for three reasons: an outgroup protagonist increases the saliency of the protagonist's perspective, it increases the protagonist's expected ignorance, and it clarifies the experimenter's intent.

The aim of this dissertation was to present a collection of studies designed to prove or rebut the OAPA in FB performance. Though the main goal was to test the account on children that fell in the sensitive age range where they first pass the FB task (but not consistently pass the task), two studies were included to test the developmental trajectory of the account with 8-to-9-year-old children an adult population. See Table 19 for a summary of the studies.

Table 19. Summary of studies.

	FB Task Type	Additional trials	ToM Tasks	Social preferences	Group	Age
1	Unexpected-contents	-	5 item scale	Exp. Preferences|IRBT	Race	4-5
2	"	-	"	Exp. Preferences	MGP	4-5
3	"	-	-	"	-	4-5
4	"	TB	-	"	MGP	4-5
5	Change-of-location	Altercentric|TB|Prank	Social Activity FB	"	"	5-6
6	"	"	"	"	"	8-9
7	"	"	"	"	"	Adults

7.1. Summary of results

Study 1 explored the pragmatic influence of manipulating the protagonist's social group membership with a real social category (race) in 4-to-5-year-olds' FB performance. Performance in the unexpected-contents FB task was framed within Wellman and Liu's (2004) five-item ToM scale. Children in the outgroup condition outperformed children in the ingroup condition in the unexpected-contents FB task as well as in the Knowledge Access task and in the overall score of the five-item scale. The group manipulation also influenced children's sequential passing of the tasks; participants in the ingroup condition were closer to the pattern found in collectivist cultures whereas children in the outgroup condition followed a completely different pattern. Implicit and explicit social preference measures did not differ between conditions and were not correlated with FB performance. *The results confirmed the prediction that the outgroup protagonist would make the task pragmatically easier for children. Crucially, while the ingroup FB performance was at-chance, the outgroup FB performance reached above-chance levels. It was concluded that the advantage resulted from a combination of three pragmatic factors: saliency, expected ignorance and experimenter's intent.*

Study 2 investigated the pragmatic influence in 4-to-5-year-olds' FB performance of manipulating the protagonist and the participant's group membership with a MGP assignment. As in Study 1, participants also completed the five-item ToM scale. The difference in performance in the unexpected-contents FB task was not significant. Nonetheless, children in the outgroup condition outperformed children in the ingroup condition in the overall five-item scale and the Real-Apparent Emotion task. The group manipulation also influenced children's difficulty response patterns with the ingroup condition following the order found in individualistic cultures and the outgroup condition following a completely new pattern. The outgroup's sequential order was mainly influenced by children's performance in the RAE task. *Improved performance in the RAE task was attributed to the competitive-prompt that was triggered by adding a threat character-trait to the outgroup individual. Two possibilities were presented to explain the non-significant between-group performance differences in the unexpected-contents task: 1) a weaker "otherness" effect created by the MGP versus Study 1's physically distinct social category; 2) a potential floor effect that would have required a stronger manipulation to bring performance to above-chance levels.*

Study 3 aimed to test the influence of the saliency of the protagonist's perspective in 4-to-5-year-olds' FB performance in isolation from any group membership information. Participants were randomly assigned to the neutral condition or the stakes condition (the narrative introduced an "urgency" element: the protagonist needed the object that should have been in the box) and completed two trials of the unexpected-contents FB task. Aside from the urgency manipulation, the task differed from Study 2's task in format (it was conducted online) and in content (the check-in question that came before the test question was removed). Participants in the stakes condition outperformed participants in the neutral condition. Overall, children's performance was better than in Study 2. *This study confirmed the prediction that increasing the saliency of the protagonist's perspective by creating an urgency narrative would enhance children's performance - making it pragmatically easier to pass the task. The overall improvement in performance as compared to the previous study was attributed to the removal of the check-in question as well as the online procedure that removed the physicality of the wrong object in the test phase.*

Study 4 explored the pragmatic influence in 4-to-5-year-olds' FB performance of combining two manipulations that increased the saliency of the target's perspective: the "otherness" booster from Study 2 and the "urgency" booster from Study 3. Children were assigned to

one team through a MGP manipulation and were randomly distributed in four conditions (ingroup neutral, ingroup stakes, outgroup neutral, outgroup stakes) and completed two unexpected-contents FB trials and one TB trial. Children's performance was at-chance in the ingroup conditions and above-chance in the outgroup conditions. Performance in the ingroup neutral condition was significantly worse than that of the outgroup neutral condition and outgroup stakes condition. The negative correlation between the FB/TB tasks was observed in the ingroup conditions (FB: at-chance /TB: above-chance) and, reversely, in the outgroup stakes condition (FB: above-chance/ TB: at-chance). *This study confirmed the booster effect of the "otherness" manipulation as performance in the outgroup conditions reached above-chance levels. There was no statistically significant difference between stakes condition. It was suggested that the interaction between the booster effects was additive, but the evidence did not provide full confirmation as performance in the outgroup stakes condition was not significantly different than performance in the ingroup stakes condition and outgroup neutral condition. The negative correlation between FB/TB that influenced each condition in an inverse manner provided evidence for the influence of group membership on the target's expected knowledge.*

Study 5 investigated the pragmatic influence on 5-to-6-year-olds' FB performance that a shared-group membership between protagonist-agent and protagonist-participant would have on the protagonist's expected knowledge/ignorance. Participants completed a MGP assignment and were divided into four conditions: ingroup protagonist-ingroup agent, ingroup protagonist-outgroup agent, outgroup protagonist-ingroup agent and outgroup protagonist-outgroup-agent. Children completed four different single trials (FB, altercentric, TB, prank) of a version of the change-of-location task that measured performance in a scalar metric (in cm) instead of a binary metric (0-incorrect / 1-correct). Children also completed four different trials of the Social Activity FB task (individual neutral activity, individual mischievous activity, group neutral activity, group mischievous activity). Children's FB bias scores were significantly greater than zero in the two conditions in which the agent and the protagonist shared group membership. Children's TB bias scores were significantly greater than zero in the four conditions. Children's performance in the altercentric and prank trials across conditions were not significantly biased. Performance in the Social Activity FB trials was above-chance except in three trials where the participant shared group membership with the protagonist. *This study confirmed that the shared-group membership between agent and protagonist negatively influenced children's FB performance as they projected the agent's knowledge onto the protagonist*

and were biased towards the wrong location. Though shared membership between protagonist and participant did not yield a significant bias score in FB trials, the results in the social FB task provided partial confirmation of the influence of the protagonist-participant shared-group membership on the protagonist's expected knowledge.

Study 6 tested whether Study 5's pragmatic manipulation of a shared-group membership between the two agents would also influence the performance of older children (8-to-9-year-olds) who would otherwise already be proficient in a standard FB task. There was no difference in FB trials between conditions. Merging the four conditions together showed that children's bias scores were significantly greater than zero in FB trials. Children exhibited a TB bias in three conditions and overall TB bias was significantly different than FB bias and prank bias, but not different than the bias exhibited in altercentric trials across conditions. Children's performance in the social FB task was above chance in all trials. *This study indicated that the two agents' group membership manipulation was not strong enough to exhibit performance differences between conditions in FB trials, but the Sandbox task was sensitive enough to be able to detect a FB bias in children that would be proficient at standard FB tasks. The significant difference between FB/TB bias scores confirmed the U-shaped developmental pattern of performance found in previous studies. The biased performance in altercentric trials was suggested to be triggered by the triviality of a real-world question that also makes TB trials hard for children at this age.*

Study 7 investigated whether the shared-group membership pragmatic manipulation of Studies 5 and 6 could influence the performance of adults who have long mastered performance at a standard FB task. As with the older children, there was no difference in FB trials between conditions. Merging the four conditions together showed that, unlike the older children, adults were not biased in FB trials. Adults exhibited a TB bias in all the conditions except the outgroup-ingroup condition. Overall TB bias was significantly different than FB bias and prank bias. Participant's performance was biased towards the wrong location in the altercentric trials in the four conditions. Adults' performance in the social FB task was above chance in all trials. *This study demonstrated that the Sandbox task was not sensitive enough to detect performance differences of the shared-group manipulation (aside from the small confirmation from the outgroup-ingroup's TB performance in which the expected ignorance of an outgroup protagonist seemed to reduce adults' bias scores). Importantly, the Sandbox task proved to be sensitive enough to detect differences between trial types. Adult's bias in TB and altercentric trials was interpreted as the result of the pragmatic*

difficulty that a trivial question poses for an individual in an experimental setting, even to an individual that has developed flexible pragmatic abilities.

7.2. Integration of results

7.2.1. Aims and confirmation of the OAPA account

The influence of the protagonist's social group membership in false-belief performance in 4-to-5-year-olds

The first experimental section in this dissertation was designed to set the stage and test how manipulating the social group membership of the protagonist in relation to that of the participants influenced ToM performance in 4-to-5-year-olds. The first study confirmed that a racial outgroup protagonist improved children's performance in the unexpected-contents FB task. Importantly, the reduced pragmatic demands of the outgroup protagonist brought children's performance to above-chance levels. The first study also confirmed that an outgroup protagonist influenced children's sequential passing of the five-item ToM scale. The second study did not confirm that a minimal outgroup protagonist enhanced children's performance, but it showed evidence that a minimal outgroup target influenced children's performance in the overall scale as well the order of difficulty in which they passed the tasks. It also showed that a threatening character trait influenced children's ability to mentalize about a target's emotions in a task where there is no conflict between the perspectives of the protagonist and participant.

This section provided evidence of OAPA's general prediction: a protagonist that does not belong to one's group will be an easier target to attribute false beliefs to.

The influence of the saliency of the protagonist's perspective in false-belief performance and how it interacts with group membership in 4-to-5-year-olds

The second experimental section was designed to address the first pragmatic factor of the OAPA: the saliency of the protagonist's perspective. Studies 3 and 4 provided confirmation that manipulating the saliency of the protagonist's perspective influenced 4-to-5-year-olds' FB performance in the unexpected-contents task. There were two mechanisms through which saliency was manipulated. The first manipulation attempted to increase the saliency by giving a narrative to the character, allowing the participant to backtrack a prior relationship between the participant and the object. The second manipulation consisted in repeating

Study 2'S MGP manipulation, that is, increasing the saliency by presenting an outgroup target. The results confirmed that children's performance was boosted by the two manipulations. Moreover, performance in the TB/FB task was negatively correlated in both conditions but in a reversed pattern.

This section provided evidence of OAPA's first pragmatic factor, the saliency of the protagonist's mental states. As predicted, increasing the saliency helped children stay focused on the perspective-tracking effort – a crucial component of passing a FB task according to the attention-focus account (Rubio-Fernández, 2017).

Moreover, the TB/FB negative correlation that affected the group conditions in an inverse manner in Study 4, provided the first evidence for the second pragmatic factor of the OAPA (the protagonist's expected ignorance); expecting the protagonist's knowledge led to an incorrect answer in the FB task and a correct answer in the TB task and expecting the protagonist's ignorance led to a correct answer in the FB task and an incorrect answer in the TB task.

The influence of the protagonist' expected ignorance in false-belief performance in 5-to-6-year-olds with a continuous measure

The third experimental section was designed to further address the second pragmatic factor of the OAPA (the protagonist's expected ignorance). The starting point for the study was that children would expect an outgroup protagonist to be less competent (derived from a character-trait stereotype) and less likely to share the participant's privileged knowledge (expectation of shared-knowledge between same-group members). The manipulation consisted in changing the social group membership of both the protagonist and the agent in a continuous change-of-location task and creating four conditions that differed in the shared-group membership between agents: agent-protagonist-participant vs. agent-protagonist vs. protagonist-participant vs. agent-participant. Study 5 demonstrated that a shared-group membership between protagonist and agent made the FB task more cognitively demanding for 5-to-6-years-olds (they exhibited a positive bias towards the wrong location). It also demonstrated that shared-group membership between protagonist-agent-participant and shared-group membership between protagonist-participant made the social FB task more demanding (exhibiting at-chance performances).

This section provided further confirmation towards the second factor of the account, the pragmatic influence of an outgroup target in the protagonist's expected ignorance. This pragmatic factor was based on the finding from Ames (2004) that when facing a dissimilar target, participants were more likely to activate a stereotype instead of a projection. It was also based on the stereotype-mindreading account (Westra, 2019), projecting that high-competence traits might make true-belief attributions more probable to the protagonist and, likewise, projecting low-competence traits might trigger a prediction of the protagonist's false-beliefs and ignorance. An outgroup protagonist made false-beliefs attributions easier because they were expected to be ignorant, and the shared-group membership created an expectation of shared-knowledge between ingroup members.

The influence of the protagonist's expected ignorance in false-belief performance with an older population that is already proficient in standard FB tasks, with a continuous measure.

The fourth and last experimental section aimed to test the developmental trajectory of the outgroup advantage in a continuous FB task with a population that had already mastered standard FB tasks. Testing 8-to-9-year-olds and adults did not show evidence for an outgroup advantage, indicating that the OAPA's predictions only applied to children in the sensitive age range.

The summation of the results of the seven studies presented in this dissertation provides confirmative evidence of the OAPA in 4-to-6-year-olds' performance in standard FB tasks. The third pragmatic factor was not explicitly tested because its effect was derived from the previous two factors. More specifically, facing a protagonist whose mental states were more salient and who was perceived as less knowledgeable would make the experimenter's question less pragmatically confusing as it would not be understood as a teaching opportunity.

7.2.2. Scope and limitations of the OAPA in relation to existing evidence in the developmental field

To my knowledge, there are two other developmental studies that have addressed the influence of the protagonist's group membership in relation to that of the participant in FB paradigms where the two perspectives are pitted against one another.[24]

Sudo and Farrar (2020) tested 3-to-5-year-olds in a between-group study where they manipulated the protagonist's social group membership through a cultural group manipulation and a MGP manipulation. Children were first administered three group affiliation measures: explicit attitude, perceived similarity, and resource allocation. Then children completed two trials of the Diverse Desires task, two trials of the FB task (one unexpected-contents trial and one change-of-location trial) and two trials of the Real-Apparent Emotion task. In the FB trials, children were asked an action-prediction question in the change-of-location task: "Where will the protagonist look for the X?" and a "think" prediction in the unexpected-contents task "What does the protagonist think is in the box?". Importantly, children were asked to justify their prediction and their justification score was added to the score of the test question. For each FB trial, children could receive a score of 2 (correct test question, correct justification), a score of 1(correct test question) or a score of zero (incorrect test question). They predicted that the cultural group manipulation would induce higher ingroup affiliation than the minimal one and, more importantly, they predicted that an outgroup target would facilitate children's performance in the DD and FB task but not in the RAE task. These predictions were based on the existing literature on conflicting perspective tasks between protagonists with adult participants. These predictions were in line with the OAPA.

The two group manipulations yielded similar group affiliation and found no between-group performance differences in DD or the RAE task. Between-group performance in the FB trials was significantly different in the older 4-year-olds and the 5-year-olds (58.62–70.40 months) who demonstrated higher ingroup affiliation, with children in the outgroup

---

[24] These studies will be presented in detail to be able to compare not only the results but also the methodology followed in each study.

condition outperforming children in the ingroup condition. Exploratory analysis showed that children's performance was worse if they demonstrated higher degrees of affiliation with the target. This effect was driven by "explicit preferences" and "resource allocation" scores and not by the "perceived similarity" scores.

The evidence provided by Sudo and Farrar (2020) is in line with the OAPA's general predictions on an improved performance with an outgroup target. Interestingly, in their study the effect was only found when the participant showed high affiliation with the target. Even more interestingly, "perceived similarity" towards the target was not related to children's FB performance. The OAPA did not specifically predict how performance would relate to group membership affiliation scores. Children in the studies conducted for this dissertation completed the Explicit Preferences Task but this forced-choice task worked more as a control measure for ingroup identification than a measure of ingroup grade of affiliation. The account's predictions were based on an automatic distinction between ingroup and outgroup members that would increase the saliency of the protagonist's perspective (in a mere membership fashion) and trigger stereotype driven character-trait attributions to the protagonists, but it did not integrate a grading system of affiliation. Considering Sudo and Farrar's (2020) findings, affiliation degrees should be incorporated in future studies of the OAPA.

Aside from OAPA's predictions, Sudo and Farrar (2020) provided new evidence on the equivalent performance between the two standard FB tasks as well as the equivalence of using an action-prediction question "look for" versus a "think" question,[25] giving support to the findings in Wellman et al.'s (2001) meta-analysis.

In the other existing study, Witt et al. (2022) tested 4-year-olds' between-group performance in a change-of-location FB task with an accent social group manipulation (first three studies) and a gender social group manipulation (last study). They predicted that children would be more likely to attribute false beliefs to the outgroup target. Their predictions were based on

---

[25] See Westra (2017a) for an analysis on the pragmatic difficulties of using the verb "think" to attribute beliefs: "The combination of the infrequency with which we overtly refer to beliefs in explanation and description and the pragmatic noisiness of 'think' makes interpreting utterances containing 'think' quite challenging for the novice speaker" (p.244).

the evidence presented by Sudo and Farrar (2020) and the adult research with tasks that involved conflicting perspectives. Their predictions were also in line with the OAPA.

In the first experiment, children were introduced to two male adult individuals wearing different color t-shirts (blue and green). One adult told a story with a native accent and the other told a story with a foreign accent. After they heard both stories, they were asked which individual had spoken in a strange manner, which individual they would want to keep narrating the story and which individual they would like to play with. Children then completed four change-of-location FB trials in which they saw a video of one of the two adult individuals placing an object in a container and leaving the room. While they were away, a female adult individual wearing a white t-shirt moved the object to the other container. A pre-recorded female voice narrated the story and emphasized that the protagonist had not seen the transfer. In each FB trial, children were asked two control questions by the experimenter, one referring to the protagonist's actions ("Where did the blue one put the banana?") and one referring to the agent's actions ("Where did the white one put the banana?"). Only when children had successfully answered both control questions, were they asked the test questions. There were two test questions, the order of which was counterbalanced across participants: a look-for question ("Where will the blue one look for the banana when he comes back?") and a "think" question ("Where does the blue one think the banana is?"). The protagonist was not back in the scene for either of the test questions. Results showed that only 56% of the participants correctly identified the outgroup target[26] (at-chance levels) and their preferences for storytelling or playmate were also not different from chance. The results did not yield a performance difference between the ingroup and outgroup conditions. Children tended to attribute fewer false beliefs to the ingroup protagonist in the "think" question, but it was not significant. Performance in the two test questions was significantly correlated. Exploratory analysis tested the between-group difference among those participants that had correctly identified the outgroup target and found that participants were more likely to attribute false-beliefs, and thus pass the task, in the outgroup condition.

---

[26] They found that identification of the target's "outgroupness" was related to the order in which they had seen the two individuals.

In Experiment 2, the researchers repeated the same procedure but presented the story-telling accent manipulation twice in alternating order. This time, 67% correctly identified the outgroup target, but preferences for storytelling and playmate continued to be at chance-levels. There were no between-group differences in performance. Experiment 3 followed the same procedure but only included the "think" test question because female participants in the previous experiment had showed more FB understanding in the action-prediction test question than male participants. Seventy-three per cent of participants correctly identified the outgroup target, they exhibited a preference towards the ingroup participant to continue the storytelling, and preference for playmate continued to be at-chance.

In Experiment 4, the category chosen for the between-group manipulation was gender; participants were familiarized with two individuals that were either an ingroup or an outgroup to them. They then completed six tasks, five from Wellman and Liu's (2004) *extended*-scale (DD, DB, KA, FB and explicit-content FB) and the change-of-location task from the previous studies. Participants ingroup preferences ranged from 65% to 77%. Children's performance was not influenced by the protagonist's group membership in any of the six tasks.

The lack of performance differences between groups (aside from the exploratory analysis in Experiment 1) is not in line with the OAPA's predictions or the evidence presented in this dissertation. The researchers attributed the null findings to a series of factors that could have influenced the manipulation in ways that were not intended: 1) the age group of the protagonists (adults) that could have been perceived as an outgroup; 2) the gender of the protagonists (male) that could have been perceived as an outgroup to female participants; 3) the weak ingroup bias generated by the accent manipulation; 4) the fact that protagonists in the first three experiments only talked in the manipulation but not in the test phase and the only way to distinguish them was the color of their t-shirt – which could have been too cognitively demanding for children. They also suggested that the outgroup advantage might only occur in a very specific age range, considering how Sudo and Farrar (2020) only found a significant effect in the older 4-year-olds and the 5-year-olds.

As it happened in Sudo and Farrar (2020), Witt et al. (2022) also provided relevant experimental evidence aside from the specific OAPA's predictions but related to the

pragmatic aspects of the FB tasks. Though they did not specifically mention whether performance in the unexpected-contents task correlated with performance in the change-of-location task in Experiment 4, I assume that it was the case because otherwise it would have been reported. In line with Sudo and Farrar (2020), they reported a correlation between the "look" and "think" test questions, and, in their case, the correlation between the questions was within the same task type.[27] Lastly, they reported participant's consistency answering the "look" and "think" test questions respectively. This last data point is important as children's consistency across multiple trials in a same testing session is not always a given (Richardson et al., 2018; Warnell & Redcay, 2019). Consistent responses in Studies 3 and 4 of this dissertation were also found, but in these studies children only completed two trials.

The integration of the findings from the above studies adds more depth to the OAPA because it outlines its strengths and limitations. In broad terms, there are three variables to take into consideration to ascertain the applicability of the OAPA: age, strength of the group membership manipulation, and measures used. In the present collection of studies, 4-to-5-year-olds (48-60 months) were influenced by the group manipulation in a standard dichotomic FB task and 5-to-6-year-olds (60-82 months) were biased in a continuous FB task. Sudo and Farrar (2020) only found a performance difference in a very specific age range (52.5-70.40 months) and Witt et al. (2022) did not find an effect in children between 47-54 months. While there is an age overlap between the participants that showed an outgroup advantage in this dissertation's studies and those in Sudo and Farrar's (2020), there is also an overlap between this dissertation's sample and Witt et al.'s (2022) sample that did not yield significative results towards an outgroup advantage. As it often happens in developmental studies, age is a variable that can generate very fine-grained predictions regarding children's performance in a task (e.g., Lane et al., 2010 found performance differences in 5-year-olds - not in younger or older children; in Seehagen et al., 2018 the manipulation affected the 4-year-olds but not the 5-year-olds). Broader predictions can only be made after observing the same effect across studies, which has also been a challenge in the developmental field (Kulke & Rakoczy, 2018). The OAPA's age of influence should be better defined through a replication and extension of the current studies.

---

[27] Sudo and Farrar (2020) asked a "look for" question in the change-of-location task and a "think" question in the unexpected-contents task.

Regarding the strength of the group membership manipulation, the review of the existing studies suggests that the OAPA can only be applicable with a social category that is salient enough to trigger social categorization processes that promote ingroup biases and that remains relevant throughout the unfolding of the task up to the test phase. In the present studies, the difference between ingroup-outgroup was salient enough that it did not require an element of affiliation to prompt an outgroup advantage. Whereas race appeared to be a very strong category because of its perceptual reinforcement, bolstering the MGP with the "mean" character-trait and the reminders (i.e., "Here's Mary, she's on the green team") also proved to be a strong categorization catalyzer. In Sudo and Farrar's (2020) study the MGP manipulation was as strong as the cultural manipulation but perhaps the targets did not immediately generate a strong ingroup bias as the enhanced FB performance was shown to apply to children that showed a higher affiliation to the target, regardless of the protagonist's ingroup or outgroup membership. In Witt et al.'s (2022) first three studies they used "accent" which is one of the most salient categories during development (Kinzler & DeJesus, 2013; Kinzler & Corriveau, 2011), so salient that it can even override race (Kinzler et al., 2009). As explained above, the authors suggested that potentially the lack of ingroup bias was derived from children's difficulty in keeping track of which group the individual belonged to. They further noted that a reminder of the protagonist's identity before each task could have made the distinction more salient (as it was done in McLoughlin & Over, 2017). However, in Witt et al.'s (2022) last study, gender – which is like race both because it is perceptual and has proven to be a very salient category for children – did not show outgroup advantages.

The third and last element to consider when attempting to extend the application of the OAPA beyond the realms of this dissertation is the measures used. Though the FB tasks have been around for 30 years, or precisely because of that, there are an infinite number of versions of the tasks and many iterations to pick and choose from. This means that detecting when the absence of a significant effect of the manipulation comes from the task used, or from the lack of power of the effect itself, is quite challenging. Just in this very small representation of outgroup advantage FB research (i.e., this dissertation's studies; Sudo & Farrar's 2020 study; and Witt et al.'s 2022 experiments) each one of us conducted a FB testing procedure that differed in numerous crucial aspects: scoring system, number of trials, wording of the question, props used, inclusion of control questions, manipulation of the

agent's group membership, and presence of the protagonist in the test phase. Though Wellman et al.'s (2001) meta-analysis showed that most variables did not influence the robustness of the FB tasks, a program that aims to detect the influence of a particular pragmatic manipulation such as the current account, needs to delimit its scope to a specific delivery of the FB tasks.

7.2.3. The OAPA in relation to evidence in adult research

Confirmative evidence for the OAPA's predictions in the older children and adults was not found. Yet it is worth bearing in mind that, as mentioned in Study 7, researchers have reported an outgroup advantage in mindreading paradigms where the perspectives are in conflict in an adult population. Savitsky et al.'s (2011) findings of a better performance in the director task with an ingroup target were interpreted under the close-communication bias account. They proposed that adults that were interacting with a friend were more likely to let their guard down and make egocentric mistakes in their directions. The close-communication bias account reinforces the OAPA's first pragmatic factor, the saliency of the target's perspective that made participants in the outgroup condition pay more attention to the protagonist's perspective, incurring in less mistakes.

The paradigmatic study for the outgroup advantage with an adult population (the one that guided Sudo and Farrar's (2020) predictions of an improved performance in the outgroup condition) is Todd et al. (2011). They proposed that "a *difference mind-set*, a cognitive orientation wherein distinctive self-referential information is activated and used as a comparison standard to draw inferences about other people could provide an efficacious route to intuiting other people's minds" (p.135). In an ingroup interaction, a similar mindset is triggered, leading to an egocentric assimilation of the other's perspectives according to one's perspectives. In an outgroup context, a difference mind-set leads participants to contrast the outgroup's perspective to their own. That is, focusing on self-other differences, improves adults' mentalizing performance. This account also reinforces the first pragmatic factor of the OAPA.

7.2.4. Learnings beyond the OAPA

The set of studies in this dissertation provided three additional findings that are beyond the scope of the OAPA and are relevant to the field of developmental ToM. First, supportive evidence was provided for the booster effect of the four variables that were found in Wellman et al.'s (2001) meta-analysis: salience, presence, participation and motive. Increasing the salience of the protagonist's perspective improved performance in Studies 3 and 4, removing the physical presence of the wrong object in the unexpected-contents task in Study 3 also improved performance, and including the participant in a deceptive action to trick the protagonist reduced the bias towards the wrong location in Studies 5, 6 and 7.

Second, the child-friendly online version of the Sandbox task proved to be a measure sensitive enough to capture bias in older children that would otherwise show a proficient performance in standard FB tasks.

Third, the use of the Sandbox task with older populations showed a developmental pattern for the puzzling negative correlation between FB/TB performance and it provided supportive evidence for the pragmatic accounts that examine the pragmatic issues that the TB task poses. While Study 5 showed that 5-to-6-year-olds exhibited an equivalent bias in FB and TB trials, studies 6 and 7 showed that 8-to-9-year-olds exhibited a larger bias in TB trials versus FB trials, and adults only exhibited a bias in TB trials. Relatedly, performance in the altercentric trials closely mirrored TB performance; whereas 5-to-6-year-olds did not exhibit a bias in altercentric trials, both the 8-to-9-year-olds and the adult sample did. Biased performance in the altercentric trials was interpreted as stemming from the same pragmatic confusing quality of the TB trials, that is, being asked a trivial, incredibly easy question.

7.3. Limitations and methodological considerations for future studies

I have reported here my attempt to prove that an outgroup protagonist is a pragmatic factor that influences performance in the FB task during the sensitive years. By comparing my experimental work to the other existing evidence in the field I have set the scope and some limitations of the applicability of my account. In this section I would like to address some

methodological aspects that further delimit the applicability of the account and should be considered in future studies.

First, the collection of studies in this dissertation were conducted with two different populations, while the first study tested Spanish participants, the other six studies tested an American population. Even though two different populations were tested, the OAPA account cannot be received as a cross-cultural account because most of the research questions were only addressed with one population, American participants. Moreover, the entire sample belonged to a Western, Educated, Industrialized, Rich and Democratic (WEIRD) society (Henrich et al., 2010)[28] and most participants were Caucasian. The focus of the present account is based on the salient distinction between one's own group membership and the group membership of the protagonist, with an ingroup protagonist eliciting a stronger ingroup identification and preferential bias. Though social categorization is a core-psychological process that transcends nationality (Rhodes & Baron, 2019), studies have shown that ingroup bias can be influenced by the social position of one group regarding the other (Kinzler et al., 2012). Thus, for instance, when using a real social category such as race, the "outgroupness" effect of a protagonist needs to be understood in the context of the social hierarchy between the participant's group and the group that is chosen to represent the "other". The outgroup advantage found in the first study using the race category cannot be generalized beyond the very particular social context in which it was found, as each racial group in each country might exhibit a diverging group dynamic.

Furthermore, specific pre-existing stereotypes related to the racial groups were not accounted for in the study. The expectation that children would stereotype the protagonist as a less competent individual stemmed from the general hypothesis that an outgroup target (regardless of group membership) triggers stereotypization for being dissimilar to oneself (Ames, 2004; Westra, 2019). Future studies should evaluate children's pre-existing stereotypes on the particular character-trait dimension that is addressed - in this case,

---

[28] Nielsen et al. 2017: "there is no universal developmental context in which children grow up, nor is there a universal environment for the human mind. To understand psychological processes, thus, it is necessary to exercise caution when generalizing beyond the specific sociocultural context at hand" (p.32).

competence - to control for any interactions between the stereotypes triggered by a particular group and the generalized mere-membership stereotypes.

The MGP manipulation technically allows for an unbiased pre-existing conception of the outgroup category. The MGP mere-membership effect suggests that an own-group positivity bias can emerge "in response to any dimension of similarity shared between two individuals even in the absence of any sort of competitive or cooperative primes" (Rhodes & Baron, 2019, p.366). As mentioned in the introduction, the origin of the own-group positivity bias might come from two opposite mechanisms; the *group*'s high value permeates to the self, or it is *one*'s high value that is transferred to the group (Dunham, 2018a). Importantly, culture might be a factor that influences whether children's mere-membership bias comes from an enhancement of the self or an integration of the group's positive status. As an example, there is evidence that interdependent or collectivist societies in which self-enhancement is considerably less established (e.g., East Asians) show ingroup favoritism to the same or even higher degree than more individual societies (Dunham, 2018a). Therefore, in future studies, the mechanism that generates the own-group positive bias in an MGP should be evaluated under the lens of the participant's culture.

Moreover, in the present studies, a competitive prime was added to the MGP to increase the ingroup-outgroup distinction. The OAPA account should also be tested with a standard MGP without the negative-valence character trait (meanness).

Second, the studies presented here analyzed children's performance in the two standard FB tasks, one was presented in their "standard"[29] forced-choice format and one was presented in a continuous measurement format. Through the present collection of studies, I referred to the important differences in the narrative set-up and the perspective-taking demands of the two FB task types, but experimental evidence of the equivalence or non-equivalence in performance of the two tasks was not provided. The initial intention was to include an unexpected-contents FB task in Study 5 to be able to compare children's performance between tasks, but it was decided not to include it to prevent any carry over effects between

---

[29] As seen in the Scope and Limitations section, the variability within the standard version is too great not to be addressed and carefully considered.

them. To aim for a generalization of the OAPA, future studies should compare children's performance in both tasks. Further, they should compare performance in the standard format of the tasks. For instance, one of the differing aspects that was presented was the physicality of the object in both tasks, and in the case of the Sandbox task, the difference is already reduced because the two possible locations are not delimited to a container; the object is not as "physically present" as it would be in a standard change-of-location task.

Third, all the studies in this dissertation included an explicit preferences task as a control for children's ingroup preferences. Considering Corriveau and Harris' (2009) halo effect account and Lane et al.'s (2013) findings on children's inferences of an informant's knowledge based on their character traits (i.e., nice and honest characters were expected to be equally knowledgeable about the content of the box as smart characters), it was expected that the positive evaluation of an ingroup would also extend to their competence evaluation. Moreover, previous studies have shown that an ingroup membership positively influences children's trust and endorsement of the individuals' competence (Kinzler et al., 2011). Nonetheless, the present studies did not include a specific question on knowledge attribution. Including such question would help further ascertain children's knowledge expectations according to group membership and would strengthen my predictions on the outgroup protagonist's expected ignorance.

Fourth, older children and adults' positive bias scores in the altercentric trials of the Sandbox task were attributed to the pragmatic confusion that these trials posed. It was suggested that the triviality of the question made participants perform similarly as they would in a TB trial. In these studies, following the example of Samuel et al. (2018a), the order of trials was maintained the same across conditions and participants. Since the focus of the studies was to examine FB bias, all participants completed the FB first. The altercentric trial was completed right after the FB by all participants. Speiger et al. (2022) found that the order in which the egocentric trials (FB in my studies) and the altercentric trials were presented influenced performance; the altercentric bias was only present when it was completed after the egocentric trial. This finding could provide support to the interpretation put forward in this dissertation, that the bias results from a pragmatic confusion rather than an altercentric interference. Nonetheless, in future studies, the order in which the trials are presented should be counterbalanced to further provide confirmation for the pragmatic interpretation.

Lastly, to my knowledge, this is the first account to manipulate the agent's group membership in relation to that of the protagonist in change-of-location tasks, creating a shared condition and a not-shared condition. In McLoughlin and Over (2017) and Todd et al. (2011) agent and protagonist always shared condition, and in Witt et al. (2022) and Sudo and Farrar (2020) the agent was neutral (a woman wearing a white t-shirt, and the character's mother respectively). Though in this dissertation I attended to the group membership of three of the partakers in a FB task, I did not manipulate the identity of the fourth partaker, the experimenter. In the first study the experimenter was an ingroup to the participants (Caucasian). In the rest of the studies, the experimenter was technically neutral to the participant's minimal group, however the way in which participants were celebrated for being in the green team and not the orange team ("Yay, you're on the green team!") could have led participants to assume the experimenter's affiliation with the green team. Pondering the weight that the experimenter's communicative action has on children's pragmatic understanding of the FB task (Helming et al., 2016), future studies should also consider the experimenter's identity in the manipulation.

7.4. Social relevance and implications

This dissertation has aimed to bring together the fields of ToM development and intergroup cognition through a pragmatic lens. Expectedly, the confirmative evidence presented here for the OAPA bears implications for both fields of study. On the cognitive side, the OAPA contributes to the pragmatic accounts that examine the context and dressing of the FB task to expose the pragmatic issues that the task presents for children. The finding that an outgroup protagonist improves children's performance in a FB task supports a more flexible understanding between competence and performance in children's mentalizing skills applied to the FB task. It does not provide support for either a nativist or an empiricist approach because it addresses a population that has already started passing FB tasks while still showing an inconsistent performance (the ½ scorers from Apperly (2012) or a performance that is not bulletproof to different presentations of the task.

On the social side, the OAPA's implications are two-fold: it poses intriguing questions for the field of intergroup cognition, and it offers a platform to develop intervention strategies to prevent children's category-based social preferences.

In Rakoczy's (2014) commentary, the researcher mentions Todd et al.'s (2011) findings and poses some of these intriguing questions:

> Does this speak in favour of the old Machiavellian idea about the origins of ToM (such that ToM arose as social-cognitive device for competition and manipulation rather than for understanding and cooperation)? Or does it merely reflect the fact that perceived difference facilitates the detection of further differences, for example, in perspectives (which is exactly what is required in standard ToM tasks)? (p.256)

While the evolutionary origins of ToM have generated diverging interpretations, the applicability of ToM in both cooperative and competitive settings is widely recognized. As introduced in the first chapter, Apperly and Butterfill (2009) explain that ToM skills need to be fast enough to provide an on-the-go guidance for competitive and cooperative activities and efficient enough to save the necessary cognitive resources for the actual task of cooperation or competition. Tsoi et al. (2021) evaluated children's ToM abilities in competitive and cooperative settings using a two-person game with stickers. In the competitive condition, the child hid a sticker under a cup - without the experimenter seeing where it went – and was then asked by the experimenter where the sticker was. If the experimenter found the sticker they kept the sticker, if they did not find it the child kept the sticker. In the cooperative condition, the child hid two stickers. If the experimenter found the stickers, no one won. If the experimenter did not find the sticker, they each got a sticker. This means that, in the two conditions, the participant needed to trick the experimenter into checking the cup that did not have stickers. The researchers found that young 4-year-olds and adults were better at tricking the experimenter in the cooperation condition. These results are particularly interesting because to achieve their goal participants needed to lie, which is an ability that would more likely be connected to a competitive goal than a cooperative one. Lee (2013) states that "lying in essence is ToM in action, because to lie and lie successfully, individuals must understand their mental state and their listener's mental states" (p.91).

The findings in this dissertation could be interpreted under the light of Rakoczy's first provocative suggestion: an outgroup target facilitates mental state reading because ToM emerged during evolution to provide the necessary tools to survive in a competitive environment. Under this interpretation, ToM's enhanced performance when facing an outgroup target should not be limited to FB measures, it should instead extend to all types of mindreading tasks. As it has been reviewed here, in mental-state attribution paradigms or tasks in which the mentalizing effort does not include conflicting perspectives, participants exhibit an advantage when facing an ingroup target.

The OAPA's implications are more in line with Rakoczy's second suggestion, the outgroup advantage stems from the differentiation between perspectives, the separation of one's experience from the other's experience. It is precisely this differentiation between perspectives that can potentially be integrated in an intervention strategy to reduce category-based social biases.

As presented in the introduction, the preschool years are crucial in the development of children's category-based social biases. Rhodes and Baron (2019) explain that even though the perceptual and conceptual social foundations of social categorization emerge quite early in development, it is during the preschool years that social categories become meaningful and a resourceful guide to predict behavior. It is also during the preschool years that children begin to integrate stereotypes and prejudices from the social interactions they have with their surroundings. Because they have just started to be exposed to cultural stereotypes and biases, Rhodes and Baron (2019) state that childhood is a "promising time to intervene" (p.372), before the biases become too entrenched in the individual's social navigation system. Skinner and Meltzoff (2019) recently reviewed all the existing research in intervention strategies to reduce intergroup bias. They identified three experiences that were consistently linked with a reduced intergroup bias during childhood: positive, cooperative intergroup contact, explicit education about prejudice and reading about or imagining contact with intergroup members.

The preschool years are also crucial in the development of ToM skills. These abilities are key for social interactions and a predictor of social ability (Astington & Jenkins, 1995; Davis-Unger & Carlson, 2008; Mizokawa & Koyasu, 2015; Slaughter et al., 2015). Interestingly, ToM has proved to be a relatively flexible set of skills that is susceptible to training (Hoyos

et al. 2020). A recent meta-analysis on studies that exposed children to training interventions to improve their ToM (Hofmann et al., 2016) showed a considerable effect size, confirming that ToM abilities can be improved by supervised experience.

The learnings from the present dissertation offer a creative idea to reduce intergroup biases while enhancing ToM skills. If imagining contact with intergroup members is a positive experience that is associated with reduced intergroup bias, and children are better at passing a false-belief task when imagining an outgroup member, a training program that presented children with ToM tasks that depicted multiple intergroup targets could potentially help children improve their perspective-taking skills whilst reducing their biases towards outgroup individuals.

Chapter 8.
Concluding remarks

The proposal presented in this dissertation is the first pragmatic account to approach the cognitive branch of Developmental Intergroup ToM. It was presented as a pragmatic endeavor because it stemmed from the premise that performance in FB tasks is not entirely dependent on children's competence but is rather entwined with many pragmatic aspects that make the task cognitively demanding and context dependent. It was framed within the cognitive branch of Developmental Intergroup ToM because it examined the cognitive implications that result from the interplay between social intergroup biases and task related cognitive biases in false belief attribution.

The learnings from the fields of intergroup cognition and developmental ToM served as the springboard for this proposal, which was named the Outgroup Advantage Pragmatic Account (OAPA). It was hypothesized that an outgroup protagonist would reduce the processing cognitive load of the FB task because it would induce a direct effect on three pragmatic factors that would influence performance: 1) it would increase the saliency of the protagonist's perspective; 2) it would create an expectation of the protagonist's ignorance; 3) it would clarify the experimenter's intent.

Seven studies were conducted to address different questions, 1) the influence that manipulating the protagonist's group membership in relation to the participant would have on children's FB performance, 2) the influence of the saliency of the protagonist's perspective on participant's performance, and 3) the influence that the expectation of the protagonist's ignorance would yield on children's bias scores. Finally, the developmental trajectory of the influence of the protagonist's expected ignorance was also explored by testing two older populations that would be expected to be proficient at the standard FB task.

The results obtained in all the studies confirm the Outgroup Advantage Pragmatic Account (OAPA): an outgroup protagonist positively influences children's performance in a FB task during the sensitive years. This account should be understood with its scope and limitations and should inspire intervention strategies to reduce intergroup bias.
The evidence provided in this dissertation confirms that the social group membership of the protagonist in relation to that of the participant's is a pragmatic factor that influences children's performance in standard FB tasks during the sensitive years. This is the first step

towards building a comprehensive view of the interaction between two social cognitive processes that play a key role in preparing the infant for a successful navigation of the social life.

Considering the weight that each of these processes might have in the development of intergroup biases, it would be worthwhile to deepen our understanding on how these processes interact and find a way to leverage the learnings from each field to build a platform to address intergroup biases at their inception.

Future research efforts should follow a systematic approach to analyze the influence of the target's social group membership in mentalizing through an exhaustive comparison of social categories that differ in nature and status. Prospective questions should start from standardized FB tasks in which the protagonist's perspective conflicts with the participant's and progressively build up to more elaborate mentalizing schemes.

Chapter 9. References

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16. https://doi.org/10.1016/S0885-2014(00)00014-9

Aberson, C. L., Healy, M., & Romero, V. (2000). Ingroup bias and self-esteem: A meta-analysis. *Personality and Social Psychology Review*, 4, 157–173.

Aboud, F. (1988). *Children and prejudice*. Oxford, England: Blackwell

Aboud, F. E., & Mitchell, F. G. (1977). Ethnic role taking: The effects of preference and self-identification. *International Journal of Psychology*, *12*(1), 1–17 https://doi.org/10.1080/00207597708247370

Aboud, F.E. (2003). The formation of in-group favoritism and out-group prejudice in young children: are they distinct attitudes? *Developmental Psychology*, 39,48–60. doi:10.1037/ 0012-1649.39.1.48

Abrams, D., Palmer, S. B., Rutland, A., Cameron, L., & Van de Vyver, J. (2014). Evaluations of and reasoning about normative and deviant ingroup and outgroup members: Development of the black sheep effect. *Developmental Psychology*, *50*(1), 258–270. https://doi.org/10.1037/a0032461

Adams, R. B., Rule, N. O., Franklin, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., … Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience*, *22*(1), 97–108. https://doi.org/10.1162/jocn.2009.21187

Adobe Inc. (2019). *Adobe Photoshop*. Retrieved from https://www.adobe.com/products/photoshop.html

Ames, D. R. (2004). Inside the mind reader's tool kit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology*, *87*(3), 340–353. https://doi.org/10.1037/0022-3514.87.3.340

Amodio, D. (2014) The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience* 15, 670–682. https://doi.org/10.1038/nrn3800

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Developmental origins of the other-race effect. *Current Directions in Psychological Science*, 22, 173–178. http://dx.doi.org/10 .1177/0963721412474459

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., & Butterfill, S. A. (2009). Do Humans Have Two Systems to Track Beliefs and Belief-Like States? *Psychological Review*, *116*(4), 953–970. https://doi.org/10.1037/a0016923

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, 35, 1311–1320. https://doi.org/10. 1037/0012-1649.35.5.1311

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Baker, E. R., Huang, R., Battista, C., & Liu, Q. (2021). Theory of mind development in impoverished U.S. children and six cross-cultural comparisons. *Journal of Applied Developmental Psychology*, *76*(August), 101314. https://doi.org/10.1016/j.appdev.2021.101314

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer Relations and the Understanding of Faux Pas: Longitudinal Evidence for Bidirectional Associations. *Child Development*, *82*(6), 1887–1905. https://doi.org/10.1111/j.1467-8624.2011.01669.x

Baron-Cohen, S., Leslie, A. M., & Frith (1985). Does the autistic child have a "theory of mind?" *Cognition*, 21, 37–46.

Baron, A.S., & Banaji, M.R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, 17, 53–58

Ben-Yami, H., Ben-Yami, M., & Ben-Yami, Y. (2019). What does the so-called False Belief Task actually check? *PsychArchives*. https://doi.org/10.23668/psycharchives.2513.

Bernstein, D. M., Thornton, W. L., & Sommerville, J. A. (2011). Theory of mind through the ages: Older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. *Experimental Aging Research*, *37*(5), 481–502. https://doi.org/10.1080/0361073X.2011.619466

Bian, L., Leslie, S.J, Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science* 355:389–91

Bian, L., Sloane, S., & Baillargeon, R. (2018). Infants expect ingroup support to override fairness when resources are limited. *PNAS* 115(11):2705–10

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, 16(3), 162–166. https://doi.org/10.1111/j.1467-8721.2007.00496.x

Birch, S. A. J., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences*, *8*(6), 255–260. https://doi.org/10.1016/j.tics.2004.04.011

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs: Research report. *Psychological Science*, *18*(5), 382–386. https://doi.org/10.1111/j.1467-9280.2007.01909.x

Braten, S. (2002). Altercentric perception by infants and adults in dialogue: Ego's virtual participation in Alter's complementary act. *Advances in Consciousness Research*, Vol. 42. Mirror neurons and the evolution of brain and language (pp. 273–294). https://doi.org/10.1075/aicr.42.20bra

Braten, S. (2004). Hominin infant decentration hypothesis: Mirror neurons system adapted to subserve mother-centered participation. *Behavioral and Brain Sciences*, 27(4), 508–509. https://doi.org/10.1017/S0140525X0427011X

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

Chalik, L., & Rhodes, M. (2014). Preschoolers Use Social Allegiances to Predict Behavior. *Journal of Cognition and Development*, *15*(1), 136–160. https://doi.org/10.1080/15248372.2012.728546

Chalik, L., & Rhodes, M. (2015). The communication of Naïve theories of the social world in parent–child conversation. *Journal of Cognition and Development*, 16(5), 719–741. https://doi.org/10.1080/15248372.2014.949722

Cimpian, A., & Markman, E. M. (2011). The Generic/Nongeneric Distinction Influences How Children Interpret New Information About Social Others. *Child Development*, *82*(2), 471–492. https://doi.org/10.1111/j.1467-8624.2010.01525.x

Coburn, P. I., Bernstein, D. M., & Begeer, S. (2015). A new paper and pencil task reveals adult false belief reasoning bias. *Psychological Research*, *79*(5), 739–749. https://doi.org/10.1007/s00426-014-0606-0

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, *12*(3), 426–437. https://doi.org/10.1111/j.1467-7687.2008.00792.x

Costello, K., & Hodson, G. (2014). Explaining dehumanization among children: The interspecies model of prejudice. *British Journal of Social Psychology*, *53*(1), 175–197. https://doi.org/10.1111/bjso.12016

Cuddy, A. J., Fiske, S. T., Kwan, V. S., Glick, P., Demoulin, S., Leyens, J. P., Bond, M. H., Croizet, J. C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H. J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., Palacios, M., … Ziegler, R. (2009). Stereotype content model across cultures: towards universal similarities and some differences. *The British journal of social psychology*, *48*(Pt 1), 1–33. https://doi.org/10.1348/014466608X314935

Cvencek, D., Kapur, M., Meltzoff, A.N. (2015). Math achievement, stereotypes, and math self-concepts among elementary-school students in Singapore. *Learning and Instruction, 39,* 1–10. https://doi.org/10.1016/j.learninstruc.2015.04.002

Davis-Unger, A. C., & Carlson, S. M. (2008). Development of teaching skills and relations to theory of mind in preschoolers. *Journal of Cognition and Development*, 9, 26–45. doi:10.1080/15248370701836584

Davoodi, T., Soley, G., Harris, P. L., & Blake, P. R. (2020). Essentialization of Social Categories Across Development in Two Cultures. *Child Development*, *91*(1), 289–306. https://doi.org/10.1111/cdev.13209

de Gracia, M. R. L., de Rosnay, M., Hawes, D., & Perez, M. V. T. (2020). Deafness and Theory of Mind Performance: Associations among Filipino Children, Adolescents, and Young Adults. *Journal of Cognition and Development*, *21*(3), 326–347. https://doi.org/10.1080/15248372.2020.1741364

De La Cerda, C., & Warnell, K. R. (2020). Young children's willingness to deceive shows in-group bias only in specific social contexts. *Journal of Experimental Child Psychology*, *198*, 104906. https://doi.org/10.1016/j.jecp.2020.104906

Deák, G. O., & Enright, B. (2006). Choose and choose again: Appearance-reality errors, pragmatics and logical ability. *Developmental Science*, *9*(3), 323–333. https://doi.org/10.1111/j.1467-7687.2006.00496.x

Deák, G.O., Ray, S.D., & Brenneman, K. (2003). Children's perseverative appearance–reality errors are related to emerg- ing language skills. *Child Development*, 74, 944–964

DeJesus, J. M., Hwang, H. G., Dautel, J. B., & Kinzler, K. D. (2017). Bilingual children's social preferences hinge on accent. *Journal of Experimental Child Psychology*, *164*, 178–191. https://doi.org/10.1016/j.jecp.2017.07.005

DeJesus, J. M., Hwang, H. G., Dautel, J. B., & Kinzler, K. D. (2018). "American = English Speaker" Before "American = White": The Development of Children's Reasoning About Nationality. *Child Development*, *89*(5), 1752–1767. https://doi.org/10.1111/cdev.12845

Devine, R. T., & Hughes, C. (2018). Family Correlates of False Belief Understanding in Early Childhood: A Meta-Analysis. *Child development*, *89*(3), 971–987. https://doi.org/10.1111/cdev.12682

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology, 52* 758-771. https://doi.org/10.1037/dev0000105

Ding, X. P., Teo, S., & Tay, C. (2021). The link between parental mental state talk and children's lying: An indirect effect via false belief understanding. *Journal of experimental child psychology*, *201*, 104990. https://doi.org/10.1016/j.jecp.2020.104990

Ding, X. P., Wellman, H. M., Wang, Y., Fu, G., & Lee, K. (2015). Theory-of-Mind Training Causes Honest Young Children to Lie. *Psychological science*, *26*(11), 1812–1821. https://doi.org/10.1177/0956797615604628

Duh, S., Paik, J. H., Miller, P. H., Gluck, S. C., Li, H., & Himelfarb, I. (2016). Theory of mind and executive function in chinese preschool children. *Developmental Psychology*, *52*(4), 582–591. https://doi.org/10.1037/a0040068

Dunham, Y. (2013) Balanced identity in the minimal groups paradigm. *PLoS One* 8, e84205

Dunham, Y. (2018a). Intergroup Cognition. *The International Encyclopedia of Anthropology*, 1–8. https://doi.org/10.1002/9781118924396.wbiea1767

Dunham, Y. (2018b). Mere Membership. *Trends in Cognitive Sciences*, *22*(9), 780–793. https://doi.org/10.1016/j.tics.2018.06.004

Dunham, Y., & Olson, K. R. (2016). Beyond Discrete Categories: Studying Multiracial, Intersex, and Transgender Children Will Strengthen Basic Developmental Science. *Journal of Cognition and Development*, *17*(4), 642–665. https://doi.org/10.1080/15248372.2016.1195388

Dunham, Y., Baron, A. S., & Banaji, M. R. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development*, 77, 1268–1281. https://doi:10.1111/j.1467-8624.2006.00933

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "Minimal" Group Affiliations in Children. *Child Development*, *82*(3), 793–811. https://doi.org/10.1111/j.1467-8624.2011.01577.x

Dunham, Y., Chen, E. E., & Banaji, M. R. (2013). Two signatures of implicit intergroup attitudes: Developmental invariance and early enculturation. *Psychological Science, 24*(6), 860–868. https://doi.org/10.1177/0956797612463081

Dunn, L. M., Dunn, L. M., & Arribas, D. (2010). *PPVT-III peabody: Test de vocabulario en imágenes* (2a ed.). Madrid: TEA.

Dunn, Lloyd M. (1997). *PPVT-III : Peabody picture vocabulary test.* Circle Pines, MN :American Guidance Service

Ellemers, N., & Haslam, S. A. (2012). Social identity theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 379–398). Sage Publications Ltd. https://doi.org/10.4135/9781446249222.n45

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

Fink, E., Begeer, S., Peterson, C. C., Slaughter, V., & de Rosnay, M. (2015). Friends, friendlessness, and the social consequences of gaining a theory of mind. *British Journal of Developmental Psychology*, *33*(1), 27–30. https://doi.org/10.1111/bjdp.12080

Flavell, J. H. (1977). *Cognitive development.* Englewood Cliffs, NJ: Prentice Hall.

Freeman, J. B., & Ambady, N. (2011). A Dynamic Interactive Theory of Person Construal. *Psychological Review*, *118*(2), 247–279. https://doi.org/10.1037/a0022327

Freeman, J. B., & Johnson, K. L. (2016). More Than Meets the Eye: Split-Second Social Perception. *Trends in cognitive sciences*, *20*(5), 362–374. https://doi.org/10.1016/j.tics.2016.03.003

Freeman, N. H., & Lacohée, H. (1995). Making explicit 3-year-olds' implicit competence with their own false beliefs. *Cognition*, *56*(1), 31–60. https://doi.org/10.1016/0010-0277(94)00654-4

Fritzley, V., & Lee, K. (2003). Do young children always say Yes to yes-no questions? A metadevelopmental study of the affirmation bias. *Child Development*, 74, 1297–1313

Gallup, G. G., Jr. (1970). Chimpanzees: self-recognition. *Science* 167, 86-87.

Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology*, *61*(3), 273–301. https://doi.org/10.1016/j.cogpsych.2010.06.001

Gelman, S.A. (2003). *The essential child: Origins of essentialism in everyday thought.* New York: Oxford University Press

Gelman, S.A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, *38(3)* 213–244.

Glenn, S. M., Johnson, K., & Parry, F. (1993). Onset of theory of mind: Methodological considerations. *Early Child Development and Care*, *86*(1), 39–51. https://doi.org/10.1080/0300443930860104

Glidden, J., D'Esterre, A., & Killen, M. (2021). Morally-relevant theory of mind mediates the relationship between group membership and moral judgments. *Cognitive Development*. https://doi.org/10.1016/j.cogdev.2020.100976

Gönültaş, S., Selçuk, B., Slaughter, V., Hunter, J. A., & Ruffman, T. (2020). The Capricious Nature of Theory of Mind: Does Mental State Understanding Depend on the Characteristics of the Target? *Child Development*, *91*(2), e280–e298. https://doi.org/10.1111/cdev.13223

Gonzalez, A. M., Steele, J. R., & Baron, A. S. (2017). Reducing Children's Implicit Racial Bias Through Exposure to Positive Out-Group Exemplars. *Child Development*, *88*(1), 123–130. https://doi.org/10.1111/cdev.12582

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59(1), 26–37.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories.* The MIT Press.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, 7(1-2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x

Green, B. F. (1956). A method of scalogram analysis using summary statistics. *Psychometrica*, 21, 79 – 88

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480. http:// dx.doi.org/10.1037/0022-3514.74.6.1464

Grice, H. P. (1975). *Logic and conversation.* In P. Cole & J. L. Morgan (Eds.), Syntax and semantics: Speech acts (Vol. 3, pp. 41–58). Academic Press

Guttman, L. (1944). A basis of scaling quantitative data. *American Sociological Review*, 9, 139 – 150.

Guttman, L. (1950). *The basis of scalogram analysis.* In S. A. Stouffer, L. Guttman, E. A. Suchman, P. A. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), Measurement and prediction (pp. 60 – 90). Princeton, NJ: Princeton University Press

Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology*, *52*, 15–23. https://doi.org/10.1016/j.jesp.2013.12.001

Han Li, P., & Koenig, M.A., (2022). The roles of group membership and social exclusion in children's testimonial learning, *Journal of Experimental Child Psychology, 216. https://doi.org/10.1016/j.jecp.2021.105342*

Haskaraca Kizilay, F.N., Proft, M., Speiger, M., Grosse Wiesmann, C., Liszkowski, U., Rakoczy, H. (2022) *Tracking altercentric and egocentric biases in a continuous false belief task.* Poster presented at BCCCD 2022, Virtual

Heller, J. (1961) *Catch-22*, (1994 – Edition). Londres, Vintage

Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, *18*(4), 167–170. https://doi.org/10.1016/j.tics.2014.01.005

Helming, K. A., Strickland, B., & Jacob, P. (2016). Solving the Puzzle about Early Belief-Ascription. *Mind and Language*, *31*(4), 438–469. https://doi.org/10.1111/mila.12114

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature.* https://doi.org/10.1038/466029a

Hetherington, C., Hendrickson, C., & Koenig, M. (2014). Reducing an in-group bias in preschool children: The impact of moral behavior. *Developmental Science*, *17*(6), 1042–1049. https://doi.org/10.1111/desc.12192

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, *344*(6190), 1243091. https://doi.org/10.1126/science.1243091

Hirschfeld, L. A. (1995). Do children have a theory of race? *Cognition*, *54*(2), 209–252. https://doi.org/10.1016/0010-0277(95)91425-R

Hofmann, S. G., Doan, S. N., Sprung, M., Wilson, A., Ebesutani, C., Andrews, L. A., .. . Harris, P. L. (2016). Training children's theory-of-mind: A meta-analysis of controlled studies. *Cognition*, 150, 200– 212. https://doi.org/10.1016/j.cognition.2016.01.006

Hofstede Insights. 2018. Country Comparison - Hofstede Insights. [online] Available at: <https://www.hofstede-insights.com/country-comparison/thailand/> [Accessed 7 May 2019].

Hoyos, C., Horton, W. S., Simms, N. K., & Gentner, D. (2020). Analogical Comparison Promotes Theory-of-Mind Development. *Cognitive Science*, *44*(9). https://doi.org/10.1111/cogs.12891

Huttenlocher, J., Newcombe, N., & Sandberg, E.H. (1994) The coding of spatial location in young children. *Cognitive Psychology*. 27, 115–147. doi:10. 1006/cogp.1994.1014

Hwang, H. G., Debnath, R., Meyer, M., Salo, V. C., Fox, N. A., & Woodward, A. (2021). Neighborhood racial demographics predict infants' neural responses to people of different races. *Developmental Science*, *24*(4), 1–10. https://doi.org/10.1111/desc.13070

Kampis, D., & Southgate, V. (2020). Altercentric Cognition: How Others Influence Our Cognitive Processing. *Trends in Cognitive Sciences*, *24*(11), 945–959. https://doi.org/10.1016/j.tics.2020.09.003

Karmakar, A., & Dogra, A. K. (2019). Assessment of Theory of Mind in Adults: Beyond False Belief Tasks. *Activitas Nervosa Superior*, *61*(3), 142–146. https://doi.org/10.1007/s41470-019-00028-1

Kawakami, K., Young, H., & Dovidio, J. F. (2002). Automatic stereotyping: Category, trait, and behavioral activations. *Personality and Social Psychology Bulletin*, *28*(1), 3–15. https://doi.org/10.1177/0146167202281001

Kawakami, K., Young, H., & Dovidio, J. F. (2002). Automatic stereotyping: Category, trait, and behavioral activations. *Personality and Social Psychology Bulletin*, *28*(1), 3–15. https://doi.org/10.1177/0146167202281001

Kelly, D.J., Quinn, P.C., Slater, A.M., Lee, K., Gibson, A., Smith, M., Ge, L., & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, 8, F31–F36.

Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology*, 26, 165–208

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

Kim, Y. S. G. (2020). Theory of mind mediates the relations of language and domain-general cognitions to discourse comprehension. *Journal of Experimental Child Psychology*, *194*, 104813. https://doi.org/10.1016/j.jecp.2020.104813

Kinzler, K. D., & DeJesus, J. M. (2013). Children's sociolinguistic evaluations of nice foreigners and mean Americans. *Developmental Psychology*, *49*(4), 655–664. https://doi.org/10.1037/a0028740

Kinzler, K. D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race?. *Cognition*, *119*(1), 1–9. https://doi.org/10.1016/j.cognition.2010.10.019

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science*, *14*(1), 106–111. https://doi.org/10.1111/j.1467-7687.2010.00965.x

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *PNAS*, *104*(30), 12577–12580. https://doi.org/10.1073/pnas.0705345104

Kinzler, K. D., Shutts, K., & Spelke, E. S. (2012). Language-based Social Preferences among Children in South Africa. *Language Learning and Development*, *8*(3), 215–232. https://doi.org/10.1080/15475441.2011.583611

Kinzler, K. D., Shutts, K., Dejesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social cognition*, *27*(4), 623–634. https://doi.org/10.1521/soco.2009.27.4.623

Kristen, S., Thoermer, C., Hofer, T., Aschersleben, G., & Sodian, B. (2006). Skalierung von "theory of Mind"-Aufgaben [Validation of the "Theory of Mind" Scale]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38(4), 186–195. https://doi.org/10.1026/0049-8637.38.4.186

Kulke, L., & Rakoczy, H. (2018). Implicit Theory of Mind – An overview of current replications and non-replications. *Data in Brief*, *16*, 101–104. https://doi.org/10.1016/j.dib.2017.11.016

Kuntoro, I. A., Peterson, C. C., & Slaughter, V. (2017). Culture, Parenting, and Children's Theory of Mind Development in Indonesia. *Journal of Cross-Cultural Psychology*, 48(9), 1389–1409. https://doi.org/10.1177/0022022117725404

Lane, J. D., Wellman, H. M., & Evans, E. M. (2010). Children's understanding of ordinary and extraordinary minds. *Child Development*, *81*(5), 1475–1489. https://doi.org/10.1111/j.1467-8624.2010.01486.x

Lane, J. D., Wellman, H. M., & Gelman, S. A. (2013). Informants' traits weigh heavily in young children's trust in testimony and in their epistemic inferences. *Child Development*, *84*(4), 1253–1268. https://doi.org/10.1111/cdev.12029

Lecce, S., & Devine, R. T. (2022). Theory of mind at school: Academic outcomes and the influence of the school context. *Infant and Child Development*, *31*(1), 1–7. https://doi.org/10.1002/icd.2274

Lecce, S., Bianco, F., Devine, R. T., Hughes, C., & Banerjee, R. (2014). Promoting theory of mind during middle childhood: A training program. *Journal of Experimental Child Psychology*, *126*, 52–67. https://doi.org/10.1016/j.jecp.2014.03.002

Lee, K. (2013). Little Liars: Development of Verbal Deception in Children. *Child Development Perspectives*, *7*(2), 91–96. https://doi.org/10.1111/cdep.12023

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind." *Trends in Cognitive Sciences*, 8, 528–533

Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief–desire reasoning as a process of selection. *Cognitive Psychology*, 50, 45–85

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). Preverbal Infants Infer Third-Party Social Relationships Based on Language. *Cognitive Science*, *41*, 622–634. https://doi.org/10.1111/cogs.12403

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The Origins of Social Categorization. *Trends in Cognitive Sciences*, *21*(7), 556–568. https://doi.org/10.1016/j.tics.2017.04.004

Linacre, J. M. (2003). Size vs. significance: Standardized chi-square fit statistics. *Rasch Measurement Transactions, 17*(1), 918.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 49(2), 433-442.

Liu, S. et al. (2015) Asian infants show preference for own-race but not other-race female faces: the role of infant caregiving arrangements. *Front. Psychol.* 6, 593

LoBue, V. & Thrasher, C. (2014). The Child Affective Facial Expression (CAFE) set. *Databrary*. Retrieved January 24, 2019 from http://doi.org/10.17910/B7301K

LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. *Emotion Science*, 5, 1532. http://doi.org/10.3389/fpsyg.2014.01532

Mahajan, N., & Wynn, K. (2012). Origins of 'us' versus 'them': prelinguistic infants prefer similar others. *Cognition*, 124, 227–233. doi:10.1016/j.cognition.2012.05.003

Mahy, C. E. V., Bernstein, D. M., Gerrard, L. D., & Atance, C. M. (2017). Testing the validity of a continuous false belief task in 3- to 7-year-old children. *Journal of Experimental Child Psychology*, *160*, 50–66. https://doi.org/10.1016/j.jecp.2017.03.010

Mair P, Hatzinger R (2007). "Extended Rasch modeling: The eRm package for the application of IRT models in R." *Journal of Statistical Software*, 20. https://www.jstatsoft.org/v20/i09.

Mandalaywala, T. M., Amodio, D. M., & Rhodes, M. (2018). Essentialism Promotes Racial Prejudice by Increasing Endorsement of Social Hierarchies. *Social Psychological and Personality Science*, *9*(4), 461–469. https://doi.org/10.1177/1948550617707020

McClung, J. S., Triki, Z., Clément, F., Bangerter, A., & Bshary, R. (2018). Endogenous oxytocin predicts helping and conversation as a function of group membership. *Proceedings of the Royal Society B: Biological Sciences*, 285(1882), 20180939. https://doi.org/10.1098/rspb.2018.0939

McLoughlin, N., & Over, H. (2017). Young Children Are More Likely to Spontaneously Attribute Mental States to Members of Their Own Group. *Psychological Science*, *28*(10), 1503–1509. https://doi.org/10.1177/0956797617710724

McLoughlin, N., Tipper, S. P., & Over, H. (2018). Young children perceive less humanness in outgroup faces. *Developmental Science*, *21*(2). https://doi.org/10.1111/desc.12539

Mills, C. M., & Elashi, F. B. (2014). Children's skepticism: Developmental and individual differences in children's ability to detect and explain distorted claims. *Journal of Experimental Child Psychology*, 124,1–17. https://doi.org/10.1016/j.jecp.2014.01.015

Mitchell, P., & H. Lacohée. 1991. Children's early understanding of false belief. *Cognition* 39: 107–127

Mizokawa, A., & Koyasu, M. (2011). Understanding false beliefs, hidden emotions, and social interactions among five- and six-year-olds. *Japanese Journal of Developmental Psychology*, 22, 168–178. doi:10.2117/ Psysoc.2007.291

Mulvey, K. L., Gönültaş, S., Herry, E., & Strelan, P. (2021). The Role of Theory of Mind, Group Membership, and Apology in Intergroup Forgiveness Among Children and Adolescents. *Journal of Experimental Psychology: General*, *151*(3), 613–627. https://doi.org/10.1037/xge0001094

Mulvey, K. L., Rizzo, M. T., & Killen, M. (2016). Challenging gender stereotypes: Theory of mind and peer group dynamics. *Developmental Science*, *19*(6), 999–1010. https://doi.org/10.1111/desc.12345

Nesdale, D., Durkin, K., Maass, A., & Griffiths, J. (2005). Group norms, threat, and children's racial prejudice. *Child Development*, *76*(3), 652–663. https://doi.org/10.1111/j.1467-8624.2005.00869.x

Newheiser, A. K., & Olson, K. R. (2012). White and Black American children's implicit intergroup bias. *Journal of Experimental Social Psychology*, 48, 264–270. doi:10.1016/ j.jesp.2011.08.011

Nguyen, L., & Frye, D. (1999). Children's theory of mind: Understanding of desire, belief and emotion with social referents. *Social Development*, *8*(1), 70–92. https://doi.org/10.1111/1467-9507.00081

Ni, Q., Fascendini, B., Shoyer, J., & Moll, H. (2022). No signs of automatic perspective-Taking or its modulation by joint attention in toddlers using an object retrieval task. *Royal Society Open Science*, *9*(8). https://doi.org/10.1098/rsos.220347

Nickerson, R. S. (1999). How we know - And sometimes misjudge - What others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737–759. https://doi.org/10.1037/0033-2909.125.6.737

Nielsen, M., Dissanayake, C., & Kashima, Y. (2003). A longitudinal investigation of self–other discrimination and the emergence of mirror self-recognition. *Infant Behavior and Development*, *26*(2), 213-226.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

O'Grady, C., Kliesch, C., Smith, K., & Scott-Phillips, T. C. (2015). The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior*, *36*(4), 313–322. https://doi.org/10.1016/j.evolhumbehav.2015.01.004

Okanda, M., & Itakura, S. (2008). Children in Asian cultures say yes to yes–no questions: Common and cultural differences between Vietnamese and Japanese children. *International Journal of Behavioral Development*, *32*, 131–136

Oktay-Gür, N., & Rakoczy, H. (2017). Children's difficulty with true belief tasks: Competence deficit or performance problem? *Cognition*, *166*, 28–41. https://doi.org/10.1016/j.cognition.2017.05.002

Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition*, *173*(January), 60–74. https://doi.org/10.1016/j.cognition.2018.01.001

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258

Otten, S. (2016). The Minimal Group Paradigm and its maximal impact in research on social categorization. *Current Opinion in Psychology*, *11*, 85–89. https://doi.org/10.1016/j.copsyc.2016.06.010

Over, H., Eggleston, A., Bell, J., & Dunham, Y. (2018). Young children seek out biased information about social groups. *Developmental Science*, *21*(3), 1–12. https://doi.org/10.1111/desc.12580

Perez-Zapata, D., Slaughter, V., & Henry, J. D. (2016). Cultural effects on mindreading. *Cognition*, *146*, 410–414. https://doi.org/10.1016/j.cognition.2015.10.018

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, 437–471. http://doi.org/10.1016/0022-0965(85)90051-7at

Peterson, C. C., & Wellman, H. M. (2009). From fancy to reason: Scaling deaf and hearing children's understanding of theory of mind and pretence. *British Journal of Developmental Psychology*, 27(2), 297–310. https://doi.org/10.1348/026151008X299728

Plötner, M., Over, H., Carpenter, M., & Tomasello, M. (2016). What is a group? Young children's perceptions of different types of groups and group entitativity. *PLoS ONE*, *11*(3), 1–14. https://doi.org/10.1371/journal.pone.0152001

Pun, A., Birch, S. A. J., & Baron, A. S. (2017). Foundations of Reasoning About Social Dominance. *Child Development Perspectives*, *11*(3), 155–160. https://doi.org/10.1111/cdep.12235

Qian, M. K., Heyman, G. D., Quinn, P. C., Messi, F. A., Fu, G., & Lee, K. (2016). Implicit racial biases in preschool children and adults from Asia and Africa. *Child Development*, 87, 285–296. http://dx.doi.org/10.1111/ cdev.12442

Quinn, P. C., Lee, K., Pascalis, O., & Tanaka, J. W. (2016). Narrowing in categorical responding to other-race face classes by infants. *Developmental Science*, 19, 362–371. http://dx.doi.org/10.1111/desc.12301

Rakoczy, H. (2014). What are the relations of thinking about groups and theory of mind? *British Journal of Developmental Psychology*, *32*(3), 255–256. https://doi.org/10.1111/bjdp.12047

Rakoczy, H., & Oktay-Gür, N. (2020). Why Do Young Children Look so Smart and Older Children Look so Dumb on True Belief Control Tasks? An Investigation of Pragmatic Performance Factors. *Journal of Cognition and Development*, *21*(2), 213–239. https://doi.org/10.1080/15248372.2019.1709467

Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests*. The University of Chicago Press, Chicago

Rhodes, M., & Baron, A. (2019). The development of Social Categorization. *Annual Review of Developmental Psychology*, Vol 1: 359–386.

Rhodes, M., & Mandalaywala, T. M. (2017). The development and developmental consequences of social essentialism. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(4). https://doi.org/10.1002/wcs.1437

Richardson, H., & Saxe, R. (2020). Development of predictive responses in theory of mind brain regions. *Developmental Science*, *23*(1), 0–3. https://doi.org/10.1111/desc.12863

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications*, *9*(1), 1–12. https://doi.org/10.1038/s41467-018-03399-2

Richter, N., Over, H. & Dunham, Y. (2016) The effects of minimal group membership on young preschoolers' social pref- erences, estimates of similarity, and behavioral attribution. *Collabra: Psychology* 2 (1)

Rizzo, M. T., & Killen, M. (2018). How social status influences our understanding of others' mental states. *Journal of Experimental Child Psychology*, *169*, 30–41. https://doi.org/10.1016/j.jecp.2017.12.008

Robbins, J. M., & Krueger, J. I. (2005). Social Projection to Ingroups and Outgroups: A Review and Meta-Analysis. *Personality and Social Psychology Review*, 9(1), 32–47. https://doi.org/10.1207/s15327957pspr0901_3

Rubio-Fernández, P. (2013). Perspective tracking in progress: Do not disturb. *Cognition*, *129*(2), 264–272. https://doi.org/10.1016/j.cognition.2013.07.005

Rubio-Fernández, P. (2017). Can We Forget What We Know in a False-Belief Task? An Investigation of the True-Belief Default. *Cognitive Science*, *41*(1), 218–241. https://doi.org/10.1111/cogs.12331

Rubio-Fernández, P. (2017). The director task: A test of Theory-of-Mind use or selective attention? *Psychonomic Bulletin and Review*, *24*(4), 1121–1128. https://doi.org/10.3758/s13423-016-1190-7

Rubio-Fernández, P. (2018). What do failed (and successful) replications with the Duplo task show? *Cognitive Development*, *48*(August), 316–320. https://doi.org/10.1016/j.cogdev.2018.07.004

Rubio-Fernández, P. (2019). Memory and inferential processes in false-belief tasks: An investigation of the unexpected-contents paradigm. *Journal of Experimental Child Psychology*, *177*(0315), 297–312. https://doi.org/10.1016/j.jecp.2018.08.011

Rubio-Fernández, P., & Geurts, B. (2013). How to Pass the False-Belief Task Before Your Fourth Birthday. *Psychological Science*, *24*(1), 27–33. https://doi.org/10.1177/0956797612447819

Rubio-Fernández, P., & Geurts, B. (2016). Don't Mention the Marble! The Role of Attentional Processes in False-Belief Tasks. *Review of Philosophy and Psychology*, *7*(4), 835–850. https://doi.org/10.1007/s13164-015-0290-z

Rubio-Fernández, P., Jara-Ettinger, J., & Gibson, E. (2017). Can processing demands explain toddlers' performance in false-belief tasks? *Proceedings of the National Academy of Sciences of the United States of America*, *114*(19), E3750. https://doi.org/10.1073/pnas.1701286114

Rubio-Fernandez, P., Long, M., Shukla, V., Bhatia, V., & Sinha, P. (2022). Visual perspective taking is not automatic in a simplified Dot task: Evidence from newly sighted children, primary school children and adults. *Neuropsychologia*, *172*(November 2021), 108256. https://doi.org/10.1016/j.neuropsychologia.2022.108256

Salter, G., & Breheny, R. (2019). Removing shared information improves 3- and 4-year-olds' performance on a change-of-location explicit false belief task. *Journal of Experimental Child Psychology*, *187*. https://doi.org/10.1016/j.jecp.2019.104665

Samuel, S., Legg, E. W., Lurz, R., & Clayton, N. S. (2018). Egocentric bias across mental and non-mental representations in the Sandbox Task. *Quarterly Journal of Experimental Psychology*, *71*(11), 2395–2410. https://doi.org/10.1177/1747021817742367

Samuel, S., Legg, E. W., Lurz, R., & Clayton, N. S. (2018). The unreliability of egocentric bias across self – other and memory – belief distinctions in the Sandbox Task. *Royal Society Open Science*, *5*(11). https://doi.org/10.1098/rsos.181355

Santiesteban, I., Catmur, C., Hopkins, S.C., Bird, G., & Heyes, C., 2014. Avatars and arrows: implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, 40 (3), 929.
Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*, *47*(1), 269–273. https://doi.org/10.1016/j.jesp.2010.09.005

Schidelko, L. P., Huemer, M., Schröder, L. M., Lueb, A. S., Perner, J., & Rakoczy, H. (2022). Why Do Children Who Solve False Belief Tasks Begin to Find True Belief Control Tasks Difficult? A Test of Pragmatic Performance Factors in Theory of Mind Tasks. *Frontiers in Psychology*, *12*(January), 1–8. https://doi.org/10.3389/fpsyg.2021.797246

Schidelko, L. P., Schünemann, B., Rakoczy, H., & Proft, M. (2021). Online Testing Yields the Same Results as Lab Testing: A Validation Study With the False Belief Task. *Frontiers in Psychology*, *12*(October), 1–7. https://doi.org/10.3389/fpsyg.2021.703238

Scott, R. M., & Baillargeon, R. (2017). Early False-Belief Understanding. *Trends in cognitive sciences*, *21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012

Scott, Rose M., Joshua C. Richman, and Renée Baillargeon. 2015. Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology*, 82. Elsevier Inc.: 32–56.

Seehagen, S., Dreier, L., & Zmyj, N. (2018). Overrated adults: 4-year-olds' false belief understanding is influenced by the believer's age. *Journal of Experimental Child Psychology*, *167*, 328–335. https://doi.org/10.1016/j.jecp.2017.11.007

Segal, S. C., Reyes, B. N., Gobin, K. C., & Moulson, M. C. (2019). Children's recognition of emotion expressed by own-race versus other-race faces. *Journal of Experimental Child Psychology*, *182*, 102–113. https://doi.org/10.1016/j.jecp.2019.01.009

Selcuk, B., Brink, K. A., Ekerim, M., & Wellman, H. M. (2018). Sequence of theory-of-mind acquisition in Turkish children from diverse social backgrounds. *Infant and Child Development*, 27. e2098 .

Setoh, P., Lee, K. J. J., Zhang, L., Qian, M. K., Quinn, P. C., Heyman, G. D., & Lee, K. (2019). Racial Categorization Predicts Implicit Racial Bias in Preschool Children. *Child Development*, *90*(1), 162–179. https://doi.org/10.1111/cdev.12851

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology*, 47, 1239–1247.

Shutts, K., & Kalish, C. W. (2021). Intuitive sociology. *Advances in Child Development and Behavior*, *61*, 335–374. https://doi.org/10.1016/bs.acdb.2021.05.004

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition*, *38*(1), 1–12. https://doi.org/10.1016/0010-0277(91)90020-5

Simpson, A. J., & Todd, A. R. (2017). Intergroup visual perspective-taking: Shared group membership impairs self-perspective inhibition but may facilitate perspective calculation. *Cognition*, *166*, 371–381. https://doi.org/10.1016/j.cognition.2017.06.003
Skinner, A. L., & Meltzoff, A. N. (2019). Childhood Experiences and Intergroup Biases among Children. *Social Issues and Policy Review*, *13*(1), 211–240. https://doi.org/10.1111/sipr.12054

Skinner, A. L., Meltzoff, A. N., & Olson, K. R. (2017). "Catching" Social Bias: Exposure to Biased Nonverbal Signals Creates Social Biases in Preschool Children. *Psychological Science*, *28*(2), 216–224. https://doi.org/10.1177/0956797616678930

Slaughter, V., & Perez-Zapata, D. (2014). Cultural Variations in the Development of Mind Reading. *Child Development Perspectives*, *8*(4), 237–241. https://doi.org/10.1111/cdep.12091

Slaughter, V., Imuta, K., Peterson, C. C., & Henry, J. D. (2015). Meta-Analysis of Theory of Mind and Peer Popularity in the Preschool and Early School Years. *Child Development*, *86*(4), 1159–1174. https://doi.org/10.1111/cdev.12372

Smogorzewska, J., Szumski, G., & Grygiel, P. (2018). Same or different? Theory of mind among children with and without disabilities. *PLoS ONE*, *13*(10), 1–20. https://doi.org/10.1371/journal.pone.0202553

Sommerville, J. A., Bernstein, D. M., & Meltzoff, A. N. (2013). Measuring beliefs in centimeters: Private knowledge biases preschoolers' and adults' representation of others' beliefs. *Child Development*, *84*(6), 1846–1854. https://doi.org/10.1111/cdev.12110

Southgate, V. (2020). Are Infants Altercentric? The Other and the Self in Early Social Cognition. *Psychological Review*, 127(4), 505–523. https://doi.org/10.1037/rev0000182

Sparks, E., Schinkel, M. G., & Moore, C. (2017). Affiliation affects generosity in young children: The roles of minimal group membership and shared interests. *Journal of Experimental Child Psychology*, *159*, 242–262. https://doi.org/10.1016/j.jecp.2017.02.007

Speiger, M. L., Langemeyer, L., Rakoczy, H., Liszkowski, U., & Grosse Wiesmann, C. (2021). *The sandbox task- a novel task to measure implicit and explicit Theory of Mind*. Poster presented at BCCCD 2021, Virtual

Speiger, M. L., Rothmaler, K., Rakoczy, H., Kizilay, F. N. H., Liszkowski, U., & Grosse Wiesmann, C. (2022). *The sandbox task-measuring implicit and explicit Theory of Mind*. Poster presented at BCCCD 2022, Virtual

Spenser, K., Bull, R., Betts, L., & Winder, B. (2020). Underpinning prosociality: Age related performance in theory of mind, empathic understanding, and moral reasoning. *Cognitive Development*, *56*(July), 100928. https://doi.org/10.1016/j.cogdev.2020.100928

Stagnaro, M. N., Dunham, Y., & Rand, D. G. (2018). Profit Versus Prejudice: Harnessing Self-Interest to Reduce In-Group Bias. *Social Psychological and Personality Science*, *9*(1), 50–58. https://doi.org/10.1177/1948550617699254

Stephenson, B., & Wicklund, R.A. (1984). The contagion of self- focus within a dyad. *Journal of Personality and Social Psychology*, 46, 163–168

Straka, B. C., Stanaland, A., Tomasello, M., & Gaither, S. E. (2021). Who can be in a group? 3-to 5-year-old children construe realistic social groups through mutual intentionality. *Cognitive Development*, *60*(June), 101097. https://doi.org/10.1016/j.cogdev.2021.101097

Sudo, M., & Farrar, J. (2020). Theory of mind understanding, but whose mind? Affiliation with the target is related to children's false belief performance. *Cognitive Development*, *54*(February). https://doi.org/10.1016/j.cogdev.2020.100869

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers Can Attribute Second-Order Beliefs. *Developmental Psychology*, *30*(3), 395–402. https://doi.org/10.1037/0012-1649.30.3.395

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. 1971. Social categorization and intergroup behaviour. *Eur J Soc Psychol*. 1: 149–178. DOI: http:// dx.doi.org/10.1002/ejsp.2420010202

Talwar, V. & Lee, K. (2008). Social and Cognitive Correlates of Children's Lying Behavior. *Child Development*, 79: 866-881. https://doi.org/10.1111/j.1467-8624.2008.01164.x

Talwar, V., & Crossman, A. M. (2012). Children's lies and their detection: Implications for child witness testimony. *Developmental Review*, 32(4), 337-359. https://doi.org/10.1016/j.dr.2012.06.004

Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology*, 43(3), 804–810. https://doi.org/10.1037/0012-1649.43.3.804

Ting, F., He, Z., & Baillargeon, R. (2019). Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(13), 6025–6034. https://doi.org/10.1073/pnas.1817849116

Todd, A. R., Hanko, K., Galinsky, A. D., & Mussweiler, T. (2011). When focusing on differences leads to similar perspectives. *Psychological Science*, *22*(1), 134–141. https://doi.org/10.1177/0956797610392929

Tomasello, M. (2009): *Why We Cooperate*. A Boston Review Book. Cambridge MA: MIT Press

Tomasello, M. (2020). The moral psychology of obligation. *Behavioral and Brain Sciences, 43*, E56. doi:10.1017/S0140525X19001742

Toyama, K. (2007). Examining theory-of-mind tasks with Japanese children: The Wellman and Liu tasks. *Japanese Journal of Educational Psychology*, 55(3), 359–369. doi:10.5926/jjep1953.55.3_359

Tsoi, L., Hamlin, J. K., Waytz, A., Baron, A. S., & Young, L. L. (2021). A cooperation advantage for theory of mind in children and adults. *Social Cognition*, *39*(1), 19–40. https://doi.org/10.1521/soco.2021.39.1.19

Tversky, B., & Hard, B.M. (2009). Embodied and disembodied cog- nition: Spatial perspective taking. *Cognition*, 110, 124–129

Vestner, T., Balsys, E., Over, H., & Cook, R. (2022). The self-consistency effect seen on the Dot Perspective Task is a product of domain-general attention cueing, not automatic perspective taking. *Cognition*, *224*(February), 105056. https://doi.org/10.1016/j.cognition.2022.105056

Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, *191*(June). https://doi.org/10.1016/j.cognition.2019.06.009

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks, *Child Development*, *75*(2), 523–541.

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

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind under-standings in Chinese children. Psychological Science, 17(12), 1075–1081. doi:10.1111/j.1467-9280.2006.01830.x

Westra, E. (2017a). Pragmatic Development and the False Belief Task. *Review of Philosophy and Psychology*, *8*(2), 235–257. https://doi.org/10.1007/s13164-016-0320-5

Westra, E. (2017b). Spontaneous mindreading: a problem for the two-systems account. *Synthese*, *194*(11), 4559–4581. https://doi.org/10.1007/s11229-016-1159-0

Westra, E. (2018). Character and theory of mind: an integrative approach. *Philosophical Studies*, *175*(5), 1217–1241. https://doi.org/10.1007/s11098-017-0908-3

Westra, E. (2019). Stereotypes, theory of mind, and the action–prediction hierarchy. *Synthese*, *196*(7), 2821–2846. https://doi.org/10.1007/s11229-017-1575-9

Westra, E. (2021). Folk personality psychology: mindreading and mindshaping in trait attribution. *Synthese*, *198*(9), 8213–8232. https://doi.org/10.1007/s11229-020-02566-7

Westra, E., & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, *158*, 165–176. https://doi.org/10.1016/j.cognition.2016.10.021

Westra, E., & Nagel, J. (2021). Mindreading in conversation. *Cognition*, *210*(January). https://doi.org/10.1016/j.cognition.2021.104618

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the Strange Stories: Revealing mentalizing impairments in autism. *Child Development*, 80, 1097–1117.

Wilks, M., Redshaw, J., Mushin, I., & Nielsen, M. (2019). A cross-cultural investigation of children's willingness to imitate prosocial and antisocial groups. *Journal of Experimental Child Psychology*, *185*, 164–175. https://doi.org/10.1016/j.jecp.2019.04.018
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128

Witt, S., Seehagen, S., & Zmyj, N. (2022). The influence of group membership on false-belief attribution in preschool children. *Journal of Experimental Child Psychology*, *222*, 105467. https://doi.org/10.1016/j.jecp.2022.105467

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA Press

Xiao, N. G., Wu, R., Quinn, P. C., Liu, S., Tummeltshammer, K. S., Kirkham, N. Z., Ge, L., Pascalis, O., & Lee, K. (2018). Infants Rely More on Gaze Cues From Own-Race Than Other-Race Adults for Learning Under Uncertainty. *Child development*, *89*(3), e229–e244. https://doi.org/10.1111/cdev.12798

Zmyj, N., & Seehagen, S. (2020). A model's competence influences false-belief attribution in preschoolers. *European Journal of Developmental Psychology*, *17*(1), 123–137. https://doi.org/10.1080/17405629.2018.1545641

Chapter 10. Appendix

Catalan translations for the five-item ToM scale (Wellman & Liu, 2004):

DD: Imagina't que és l'hora de berenar i aquest nen pot triar entre dues coses. Pot triar o [galetes] o [pastanagues]. A tu què t'agradaria més? És una bona tria! Però a aquest nen li agraden molt les [galetes]. No li agraden les [pastanagues]. I ara li toca menjar. Quin berenar triarà aquest nen? (Assenyalar dreta o esquerra). Versió B: palets de pa o tomàquets petits.

DB: Imagina't que aquesta nena està buscant el seu gat. El seu gat pot estar darrere els [arbusts] o dins el [garatge]. Tu on creus que està el gat? Vale, bona tria! Però aquesta nena es pensa que està als [arbusts]. Després, quan aquesta nena vagi a buscar el gat on el buscarà? (Assenyalar dreta o esquerra). Versió B: Sofà o llibreria.

KA: Mira aquesta capsa. Saps què hi ha dins la capsa? Què creus que hi ha dins la capsa? (Obrir la capsa) A veure... hi ha....[un gos]! (Tancar la capsa.) Vale, t'enrecordes de què hi ha dins la capsa? Imagina't que ara arriba aquesta nena. Aquesta nena mai ha vist la capsa. Sap què hi ha aquesta nena dins la capsa?

FB: Mira aquesta capsa de [tirites]. Què creus que hi ha dins la capsa de [tirites]? (Obrir la capsa) A veure... hi ha.... [plastidecors]! (Tancar la capsa.) Vale, t'enrecordes de què hi ha dins la capsa? Imagina't que ara arriba aquesta nena. Aquesta nena mai ha vist la capsa. Què pensa que hi ha dins la capsa, [tirites] o [plastidecors]?

RAE: Ara t'explicaré la història d'un nen (foto d'un nen per darrere). Et preguntaré com creus que el nen es sent per dins i la cara que fa. Pot ser que realment se senti d'una manera però que posi una cara diferent. O pot ser que es senti de la mateixa manera que la cara que fa. (anar ensenyant el paper). Aquesta és la història d'aquest nen. Els amics d'aquest nen jugaven junts i feien broma. Un dels nens més grans es va burlar d'ell i tots van riure. Tothom pensava que era molt divertit, menys ell. Però ell no volia que els altres veiessin com es sentia perquè sinó encara es burlarien més d'ell. Per tant ell va intentar amagar com es sentia. Què van fer els altres nens quan el nen més gran es va burlar d'ell? (van riure) Què farien els nens si veiessin com estava ell? (es burlarien més). Com creus que es sentia ell quan tothom reia? Es sentia content, normal o trist? Quina cara intentava fer cap a fora quan tothom reia? Content, normal o trist?

The end.