



## Research Article

**Cite this article:** Feng, J., Cho, S., & Luk, G. (2023). Assessing Theory of Mind in bilinguals: A scoping review on tasks and study designs. *Bilingualism: Language and Cognition*, 1–15. <https://doi.org/10.1017/S1366728923000585>

Received: 6 February 2023  
Revised: 19 June 2023  
Accepted: 7 July 2023

### Keywords:

bilingualism; Theory of Mind; adults; children; scoping review

### Corresponding author:

Justin Feng,  
[justin.feng@mail.mcgill.ca](mailto:justin.feng@mail.mcgill.ca)

### Abstract

Previous developmental studies reported bilinguals' Theory of Mind (ToM; the ability to take on another's perspective) develops differently than monolinguals. We conducted a scoping review to evaluate how researchers assess bilinguals' ToM and whether they characterize bilinguals' lived experiences. We analyzed 53 publications examining ToM in bilinguals, with most papers studying children ( $n = 42$ ; 79%). We identified 96 different tasks used across these 53 papers. The most common are 46 (48%) cases of the false-belief task, a cognitive-focused task using story vignettes. Few tasks target other types of ToM, such as ToM in social settings or taking others' emotional perspectives. Furthermore, only half of the papers reported language history ( $n = 28$ , 53%) and exposure ( $n = 25$ , 47%), limiting the inferrability of ToM and language experiences. Expanding how we study ToM in bilinguals will improve our understanding of the intersection of bilingualism and ToM.

### Introduction

Theory of Mind (ToM) is the ability to take on the perspective of other individuals; it is our capacity to infer others' intentions, beliefs, and emotions (Premack & Woodruff, 1978). The concept is at times associated with terminologies like mindreading (Gallese & Goldman, 1998), mentalizing (Frith & Frith, 2006), and folk psychology (Stich & Ravenscroft, 1994). ToM centralizes on the idea of perspectives – our egocentric views may or may not differ from those of another individual due to individualized knowledge, beliefs, emotions, and desires. Developmental research has shown that by the age of four, children would have acquired ToM (Wellman et al., 2001), reflected in their ability to pass laboratory ToM tasks designed for children. Early achievement of ToM is an essential developmental milestone, extending beyond social competencies, showing a positive correlation with school achievement (Smogorzewska et al., 2022), moral reasoning (Smetana et al., 2012), and empathy (Bensalah et al., 2016).

Cognitive representation of ToM is abstract and has been shown to be strongly related to language (Astonington & Baird, 2005; de Villiers & de Villiers, 2014; Milligan et al., 2007; Watson et al., 2001), particularly in language's utility to convey our thoughts and infer others' thoughts. In the English language, words like “think” and “believe” represent knowledge, feelings, and perception. ToM competence in children is therefore contingent on language proficiency, as revealed in the strong correlations between ToM performance and language abilities (de Villiers & Pyers, 2002; Pyers & Senghas, 2009). Telling lies (Lee & Imuta, 2021) or understanding irony (Filippova & Astonington, 2008) rely on language comprehension and the ability to infer the interlocutor's knowledge. Given the strong association between ToM and language abilities, researchers have also investigated whether diverse language experiences modulate ToM, such as comparing bilinguals or multilinguals to those who are monolingual. Here, adopting Grosjean's (2013) broad definition, we refer to BILINGUALS as those who speak and use two or more languages on a daily basis.

Studying bilingual consequences on ToM requires tasks relevant and meaningful to the target population. Traditionally, ToM has been studied using false-belief tasks that traced back to Wimmer and Perner (1983). Understanding false belief is the recognition that others can believe in things that are not true and conflict with the current reality (Lohmann & Tomasello, 2004; Wellman et al., 2001). For example, the Sally-Anne Task is a standard false-belief measurement (Baron-Cohen et al., 1985). In this task, children listen to or read a vignette where one character, Sally, puts a marble in her basket and then leaves the room. Anne enters the room and moves the marble from Sally's basket to her own basket. Then Sally returns. The participants are asked: “Where is the marble really?”, “Where was the marble in the beginning?” and “Where will Sally look for the marble?”. The last question serves as the false-belief question and requires the separation between the child participant's egocentric

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



This article has earned badges for transparent research practices: Open Materials. For details see the Data Availability Statement.

perspective and Sally's perspective. Therefore, the correct answer would be to indicate that Sally would look for the marble in her own basket as she is unaware of Anne's actions.

Recent evidence, however, suggests that tasks like the Sally-Anne Task may measure more than ToM and have large task variability. Warnell and Redcay (2019) used a battery of ToM measurements, such as False-belief, Appearance-Reality<sup>1</sup>, and Faux Pas<sup>2</sup>, on children and adults. They found that all age groups had a wide range of responses across these tasks showing little convergent validity across tasks, i.e., performance in one ToM task does not predict performance in the other. Additionally, Quesque and Rossetti (2020) critically reviewed a set of ToM tasks and concluded that ToM must satisfy two criteria: NONMERGING, distinction between one's own beliefs from others, and MENTALIZING, competency relevant to a high level of mental representation of knowledge rather than perceptual abilities such as attention or visual judgment. Based on these criteria and theories concerning ToM, we identified four issues in the current literature examining bilingualism and ToM. We use the following four points to guide our scoping review to address the empirical basis of bilingualism and ToM.

First, the ToM literature has focused on children, yet ToM continues to develop into adulthood (Dumontheil et al., 2010) in which the assumption of adults "mastering" ToM remains questionable as they frequently make mistakes when inferring others' beliefs and perspectives (Keysar et al., 2003). Apperly et al. (2009) further indicated the importance of using adult studies to inform ToM development and its links with executive functioning and language comprehension: both skills had been shown to demonstrate differential performances in bilinguals and monolinguals. However, simply focusing on group differences undermines the importance of attending to language experience as a social and cognitive experience (Luk, 2022; Rothman et al., 2022). The current developmental ToM tasks are not designed for adults, making it challenging to assess ToM in adults. Therefore, in this scoping review, we examine ToM tasks in studies involving bilingual children and adults.

Additionally, the disciplinary foci on ToM have been divergent in the current literature. For example, neuroimaging studies revealed a differentiation in ToM, specifically cognitive ToM, inferring beliefs and intentions, and affective ToM, inferring emotions and feelings (Poletti et al., 2012; Shamay-Tsoory et al., 2005, 2010). According to Shamay-Tsoory et al.'s (2010) model, cognitive ToM is a prerequisite for affective ToM, indicating that affective ToM reflects more advanced ToM development. The choice of ToM tasks, therefore, should consider and distinguish cognitive and affective ToM. For instance, the Yoni task assesses the participant's judgement of eye gaze and facial expressions along with both a verbal cognitive (e.g., think) or affective (e.g., love) comment (Shamay-Tsoory & Aharon-Peretz, 2007). In monolinguals, while the language itself mediates both types of ToM (Bigelow et al., 2021), Cassetta et al. (2018) found that executive functioning predicted cognitive ToM, whereas verbal IQ predicted affective ToM, suggesting the two types of ToM are dissociable. Relatedly, Han and Lee (2013) examined this differentiation in bilinguals, showing that perhaps bilinguals performed differently in affective ToM, not cognitive ToM. Therefore, we document whether the ToM tasks focused on cognitive or affective ToM.

In a similar vein, many ToM tasks do not probe into the understanding of social interactions. For instance, the aforementioned Sally and Anne task is generally a cognitive ToM task as the questions focus on Sally's perspective, a character in a story

situated in a social situation that sets up the scenario but not serving as the focus of the task. Participants express their understanding of ToM only through the forced perspective of a story vignette. They do not interact with another person, nor do they consider how to use their own ToM. Since ToM, like bilingualism, involves social interactions, the circumstantial use of tasks like Sally and Anne captures PERCEPTUAL or COGNITIVE perspective-taking rather than SOCIAL perspective-taking in ToM.

Theoretical frameworks such as the Adaptive Control Hypothesis (Green & Abutalebi, 2013) indicate the types of interactional contexts where language and conversation occur among bilinguals. Within this framework, the use of languages requires the consideration of others' available languages to allow fluid switching and successful interactions. This consideration may depend on ToM and varies based on the context. Therefore, when evaluating ToM in bilinguals, a social lens reflects the social context in which multiple languages are used in bilinguals, allowing the identification of converging processes between multilingual experiences and social interactions. Alkire et al. (2022) addressed the methodology limitation by focusing on CONVERSATIONAL ToM, which is using ToM in a conversation. They indicated that the common laboratory tasks do not emulate everyday social interactions, which prompted their novel design. Hence, we also ask whether studies in the literature make a similar differentiation between different aspects of ToM, particularly ones that involve a situational setting with social interactions.

Another source of variation in findings concerns the participants' diverse language and social experiences. The ever-shifting understanding of bilingualism suggests the need for careful characterization of the sample: onset ages when language acquisition occurs, school language use, home contexts and exposure. Hence, we take on an approach similar to "bilingual phenotyping" in this review, which broadly argues that research in bilingualism should be grounded on how well we characterize bilingual participants (Navarro-Torres et al., 2021). It is meant to avoid perpetuating the assumed homogeneity within bilingual groups in research (Surrain & Luk, 2019). The heterogeneity of bilingual experiences should be documented to identify specific experiences relevant to ToM. Here, we do not indicate there is a causal relationship between bilingualism and ToM. We aim to identify correlates relevant to both ToM and bilingualism in this scoping review.

This scoping review assessed ToM tasks used in the literature on bilingual children and adults. We also considered the constructs studied in ToM, whether the tasks focus on cognitive or affective, and whether the tasks situate in social settings. Lastly, since both bilingualism and ToM are social constructs, we investigated how the social and linguistic characteristics of the multilingual populations were documented in the ToM literature.

## Methods

### Search protocol and databases

Following the PRISMA framework (see Figure 1, Tricco et al., 2018), we searched three databases chosen to encompass a broad range of interdisciplinary literature: SCOPUS, PsycINFO (1806-Ovid), and Medline (Ovid) 1946. The literature search was conducted on January 24th, 2022. Considering that each database uses different categorizations and terminologies, we adapted different search terms for SCOPUS, PsycINFO, and Medline (see Table S1 for database-specific search terms). As discussed in the literature review, ToM is a complex phenomenon, referred to in

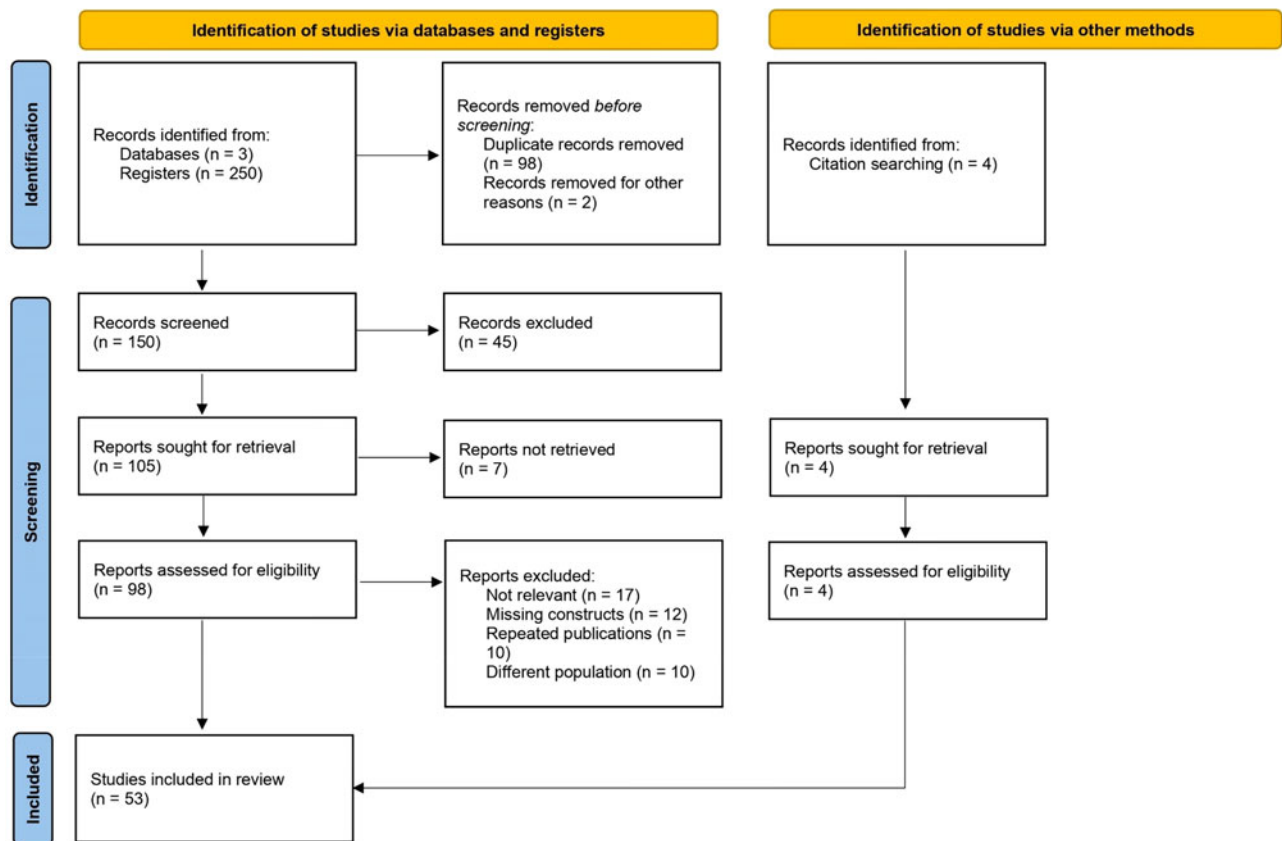


Figure 1. PRISMA Flowchart to Identify Studies to be Included in the Analysis.

the literature as mentalizing, mind-reading, perspective-taking, social intelligence, and folk psychology (Astington & Baird, 2005; Quesque & Rossetti, 2020). Where feasible, these terms were part of the search to ensure a broad scope of the literature is incorporated. Furthermore, we conducted a manual search to include publications that the three databases missed but were discussed in the literature review of papers in the database. The database and supplementary materials are available on OSF (<https://osf.io/r96ju/>).

The eligibility criteria in Table 1 have been designed to be open and broad to examine the literature comprehensively. Any paper studying bilinguals that included ToM as a construct was included, irrespective of whether ToM was an outcome variable. No time frame was established. These broad criteria permitted a more holistic review of how bilingualism and ToM are measured in tandem. Publications that did not study bilinguals and perspective-taking/ToM together that passed through the initial filters were first

excluded. Since this review examines study designs, studies must include a methods section. We kept dissertations as they tend to involve an extensive methods section. However, if the dissertation was published as an empirical paper, they were excluded from the final set with only the peer-reviewed publication included to avoid duplicated results. We only included studies on typically developing children, given that language and ToM reveals different developmental trajectories in atypical populations (Schick et al., 2007). We excluded studies using sign language(s) and those involving children with disabilities. Lastly, papers adopting longitudinal designs focused on the effects of intervention or training were also excluded due to the focus on documenting behavioral change.

### Selection of sources and data charting process

In the initial screening, duplicates were removed. Then, titles and abstracts were read by the first author to assess the publications

Table 1. Summary of Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
<ul style="list-style-type: none"> <li>Papers that examined bilinguals, including second-language learners AND</li> <li>Papers that examined perspective-taking in all sense of the term, including spatial reasoning, mentalizing, role-taking</li> </ul>	<ul style="list-style-type: none"> <li>Papers that did not include bilingual or multilinguals</li> <li>Papers that did not study any form of perspective-taking/ToM</li> <li>Papers with only atypically developing children or adults</li> <li>Papers that did not include a methods section</li> <li>Papers that looked at interventions such as developing ToM, and empathy in a target population</li> <li>Papers or dissertations that used the same dataset and methodology as an empirical paper</li> </ul>

based on the exclusion criteria. After excluding articles, the first two authors read the papers to determine eligibility. Information about each study was charted systematically: (a) general participants' demographics; (b) whether the study looked at children or adults; (c) study objective and context; (d) ToM tasks used; (e) bilingual characterization; and (f) relevant results. The first and second authors worked on coding the papers independently. Any unresolved issue resulted in both authors revisiting the paper and jointly recoding the paper.

For the ToM tasks, we created a coding scheme adapting from Yeh et al. (2021) on adults with schizophrenia. Yeh et al.'s scheme discussed the ToM concepts of each task, the modality of the tasks, and the mode of response. They also distinguished cognitive and affective ToM and examined how the tasks were scored. In this paper, we evaluated how the participants were questioned, how they responded, and in what language the exchange took place. The documentation process was summarized in Table 2. We examined and coded administration modality, testing format, response modality and format, the language of task administration, other behavioral correlates (i.e., executive functions, metalinguistic awareness, and socioeconomic status), and whether and how cognitive and affective ToM was assessed. In this case, if the ToM tasks in the study involved emotion, it is considered as affective ToM.

To document bilingual characterization, we isolated how bilinguals and monolinguals (when applicable) were defined in each paper. The coding was adapted from Surrain and Luk (2019). Table 3 includes a description of the qualitative coding scheme. Specifically, we examined: (a) how groups were determined; (b) language proficiency; (c) language history; (d) language exposure; and (e) language context (sociolinguistic context). While these are not the only factors that are part of the bilingual experience, these were selected due to their prevalence in bilingual research (Anderson et al., 2020; Byers-Heinlein et al., 2019; Surrain & Luk, 2019). Finally, the coding was completed according to the methodology sections in each study and, where applicable, questionnaires or interviews in the appendix or supplementary materials.

## Results

### Papers selected

A total of 250 papers were initially identified from the three databases. After removing 100 duplicates, the remaining 150 papers were screened to ensure they included bilinguals and studied ToM. We then read the abstracts and excluded 45 papers, leaving 105 papers for the full review. We removed 56 of the 105 publications according to the exclusion criteria and retrieval issues. Of the removed articles, 4 were dissertations that were published. Four additional papers were identified from a manual search (Bialystok & Senman, 2004; Farhadian et al., 2010; Han & Lee, 2013; Pearson, 2013). The final database included 53 papers, of which 8 were dissertations.

### General characteristics

The 53 papers, published between 1982 and 2022, had a total sample size of 5,304 participants, ranging from 24 to 317, with an age span from 2 years to 73 years. Thirty-four studies had fewer than 100 participants, 15 had between 100-200 participants, and 4 had over 200 participants. The majority of the studies looked at children ( $n = 42$ , 79%), with the remaining studies reporting ToM in adults ( $n = 10$ ; 19%), and one study included both children and adults. In

**Table 2.** Coding Scheme for ToM Tasks

Features	Description and Coding Scheme
Administration modality	Were the tasks administered verbally or nonverbally (low verbal task, such as when minimal words were used, were classified as nonverbal)? 1 = Verbal Task 2 = Nonverbal Task 3 = Both (both were included in the same study due to different tasks)
Testing format	More detail on the task format. <ul style="list-style-type: none"> <li>• Computer = Task administered digitally</li> <li>• Direction = Step-by-step directions given</li> <li>• Objects = Tasks that use a physical object, normally for spatial perspective-taking tasks, or to supplement stories</li> <li>• Pictures = A set of pictures is given</li> <li>• Reading = Reading a story/vignette of some</li> <li>• Standardized tests = Standardized as in there is a normed standard score</li> <li>• Stories = A story vignette is used</li> <li>• Verbal questions = Question was asked verbally</li> <li>• Video = Video used</li> <li>• Written = A writing task</li> </ul>
Response modality	Did the participants respond verbally or nonverbally? 1 = Verbal Response 2 = Nonverbal Response 3 = Both
Response format	More detail on the response format: <ul style="list-style-type: none"> <li>• Eye-tracking = Uses eye-tracking technology</li> <li>• Language switching = Requires language switch to match the speaker's language</li> <li>• Open-ended response = The question requires an explanation</li> <li>• MC (pictures) = Multiple choice, using pictures allowing pointing nonverbally</li> <li>• MC (text) = Multiple choice, but with text choices</li> <li>• Story-retell = Requires retelling of a story, high verbal component</li> <li>• Verbal response = Basic response to a specific question, verbal</li> <li>• Written = Written response</li> </ul>
Language of task	What language were the tasks administered? L1 = First language L2 = Second language B = Administered in both languages C = Participants gets to choose D = Depends on the participants, tasks are administered in one language (usually English) but whether it is L1 or L2 differs from participant to participant Any item marked with * is unclear in the study.
Behavioral correlates	Were executive functioning, metalinguistic awareness, and socioeconomic statuses collected, either groups are matched, or these are considered in their statistical models? 0 = No 1 = Yes
Cognitive or affective ToM	Were the tasks measuring cognitive or affective Theory of Mind? C = Cognitive A = Affective B = Both in a single task D = Both in separate/different tasks

**Table 3.** Coding Scheme for Bilingual Characterization

Features	Description and Coding Scheme
Classification	How were monolinguals and/or bilinguals classified (categorical)? 0 = unclear 1 = home and school language 2 = proficiency 3 = self-report questionnaire (self, parent, or teachers) 4 = questionnaire (determined by researchers) 5 = other
Questionnaire	If a questionnaire was used to characterize the participants, which questionnaire did they use? 0 = Unclear or unspecified 1 = Original questionnaire (either unnamed or created for the study), <i>Name of the questionnaire included</i>
Language proficiency	How are monolingual and bilingual proficiency assessed? Note that if both are given, coding defaults to formal assessment as this would normally be included in the statistical analyses (unless stated otherwise) 0 = Not assessed 1 = Formally assessed 2 = Self-report (self or parent)  If language was formally assessed, was it for one language, or both? (Code goes in conjunction with above) L1 = First language L2 = Second language B = Both
Language history	Do they report features such as the age of acquiring a certain language and the order of language? 0 = No 1 = Yes
Language exposure	Do they report the type and/or amount of language exposure? 0 = No 1 = Yes
Language context	Whether the sociolinguistic context of language use is indicated. This includes home and school contexts. We also assessed if they considered at a cultural level. Multiple responses were allowed. 0 = Not reported 1 = Cultural (depicts language use at a societal level) 2 = Home (home language use) 3 = School

terms of study designs, the majority adopted between-group comparisons on monolinguals and bilinguals ( $n = 40$ ; 76%). One study (Díaz & Farrar, 2018b) followed a longitudinal design but did not use any interventions or training. All studies included in the scoping review are reported in Table S3 and S4 in the supplementary material and are also available on OSF.

## Theory of mind tasks

### Type of ToM tasks

The first objective of this scoping review was to document and conduct a frequency count of ToM tasks used. Upon analyzing the data, several studies created a variation different from the

original procedure, making it difficult to assess frequency as they could not be considered as unique categorical counts. For example, the Director Task can be delivered in-person (Fan et al., 2015), computerized (Lorge & Katsos, 2019), or without a director (Navarro & Conway, 2021). Counting these tasks as the same Director Task is inappropriate. Therefore, we adopted Diaz's (2021) classification and combined it with the type of ToM tasks outlined by Wellman and Liu which included the diverse desires, diverse beliefs, knowledge access, contents false belief, explicit false belief, belief emotion, and real-apparent emotion (Wellman & Liu, 2004, p. 531). We also took into the considerations raised in Quesque and Rosetti (2020; see page 6 on the inclusion of mentalizing and nonmerging criteria). Table 4 reports the coding scheme with 11 types of ToM tasks. We counted that approximately 96 different tasks were used across the 53 publications<sup>3</sup>. Some studies used multiple tasks resulting in more tasks than studies. For instance, we identified ten cases of the unexpected-contents task, with at least seven separate sources of the unexpected-contents task cited<sup>4</sup>. Notably, there were 46 cases (48% when considering all 96 tasks) of false-belief tasks, followed by 11 cases (15%) of perspective-taking tasks.

In Figure 2, we present a visual representation of two weights ranging from -5 to +5 for each of the eleven types of ToM tasks across two dimensions: cognitive/affective and non-social/social. Then we translate the frequency of occurrence in tasks as they pertain to children or adults in the plot, making the size of each circle proportional to the total frequency counts of tasks, with segments pertaining to whether the task was administered for children or adults. We emphasize here that these two planes are not meant to be on an ordinal or continuous scale but a representation showing the distribution of the tasks along the cognitive-affective focus and whether the tasks involve social/non-social settings. We weighed the tasks on each dimension independently. For example, on the non-social/social dimension, weights are a categorical representation of probability, where -5 to -1 would be tasks involving LESS social interactions or circumstances and +1 to +5 for tasks that include MORE social elements (i.e., interactions with or consideration of a direct person). This entire process was first completed by the first author and later discussed with the second and third authors.

Our evaluation of the cognitive-affective dimension was contingent on whether the task assesses emotions. An example of cognitive tasks is the visual-spatial perspective-taking task used in Gorrell et al. (1982), which only involves the need to rotate objects correctly with no reference to another agent's emotions and affective states. Belief emotion tasks (tasks adapted from Wellman & Liu, 2004) are an example of an affective task. Along the social/non-social dimension, we weighed tasks on consideration of another individual's perspective in a social situation. For example, a language-switching task would require perceiving the other person's language (see Cheung et al., 2010 for an example). Non-social tasks, such as appearance-reality tasks, would be the opposite, seldom involving another direct perspective. Instead, appearance-reality tasks involve assessing objects or looking at third-person vignettes, as observed in Berguno and Bowler (2004). In cases where multiple types of ToM were assessed, such as the Theory of Mind Inventory (ToMI-2; Hutchins et al., 2012; used in Kim et al., 2021), we gave the weight of zero in both dimensions.

As two weights were given to each type of task, certain tasks can be weighted more on one dimension but not the other. For instance, scene-description tasks are primarily cognitive tasks

**Table 4.** Coding Scheme for the Type of Tasks

Type of ToM Task	Description of Classification and Sample Reference
AR - Appearance Reality	Distinguishes the difference between the appearance and true states (Flavell et al., 1983).
BE - Belief Emotion	A combination of belief emotion and real-apparent emotion tasks as defined by Wellman and Liu (2004). These tasks emphasize judging emotional discrepancies.
DD - Diverse Desire and Belief	A combination of diverse desires and diverse beliefs as defined by Wellman and Liu (2004). Both focus on the understanding that others have different desires and beliefs <b>in contrast to their own</b> .
FB - False Belief	False-belief tasks require the acknowledgement that another party can believe and act on things that are not true. Examples include unexpected contents, unexpected transfers, and other similar tasks (Wellman & Liu, 2004; Wimmer & Perner, 1983).
FP - Faux Pas	Focus on social mistakes and things that should not have been said in certain social interactions (Stone et al., 1998).
KA - Knowledge access	Perspective based on whether the other individual has access to certain knowledge such as visually seeing something, defined by Wellman and Liu (2004). These are not considered false-belief due to not assessing that something not true.
LS - Language Switching	Explicitly requires one to switch to the language of an interlocutor to complete the task successfully.
PT - Perspective Taking	Take on a non-visual perspective of another individual. It differs from other constructs (i.e., DD) in that this type of ToM emphasizes on evaluating another person's beliefs and knowledge <b>without explicit consideration of one's own</b> .
SD - Scene Description	Describe or follow a description of a scene that requires a complex and active perspective shift, such as in the Director Task.
VS - Visuospatial	A purely visuospatial-based rotation of objects or perspectives. Does not involve mental states or emotions.
MT - Multiple	Measures multiple different aspects and types of ToM (the ones within this list) in the same task, usually seen in standardized assessments

requiring the participant to assess the visual perspective of the director. However, since the presence and instruction from a director are necessary for this task, participants need to engage the director, a social element. This task also exemplifies the complexity of coding these tasks and how the coding should not be considered ordinal or continuous. As mentioned previously, the Director Task can also be done on a computer screen. Using this modality loses the social dimension as it becomes similar to reading a vignette (i.e., false belief tasks). Based on the weights we provided (see supplemental material), it is possible to interpret the types of ToM tasks as: (1) social/affective, (2) social/cognitive, (3) non-social/cognitive, and (4) non-social/affective. In Figure 2, we noted that only 21 tasks (22%) are in the social/affective quadrant. We observe a heavy representation of non-social/cognitive ToM tasks in the literature, with 60 (63%) tasks in this quadrant. Scene description and diverse desires were the only two types of ToM tasks outside the non-social/cognitive and social/affective quadrants.

Figure 2 also portrays the limited number of studies on adult ToM, with only 11 out of the 53 publications including adults. We observed seven of the 11 types of ToM tasks being used in the adult studies (a total of 13 tasks across the 11 studies), two cases of both perspective-taking and visuospatial tasks and three cases of false-belief and scenario description tasks. The tasks used for adults are more complex than the children's, except for the three cases of false-belief tasks. Two were from the same author in a neuroimaging study (Kobayashi et al., 2007, 2008). The third study by Rubio-Fernández and Glucksberg (2012) noted the concern of using a false-belief task with adults, though eye movements and reaction times were used to ensure variability. The Director Task that uses scenario descriptions was predominantly used with adults, with only one case used with children (Fan et al., 2015).

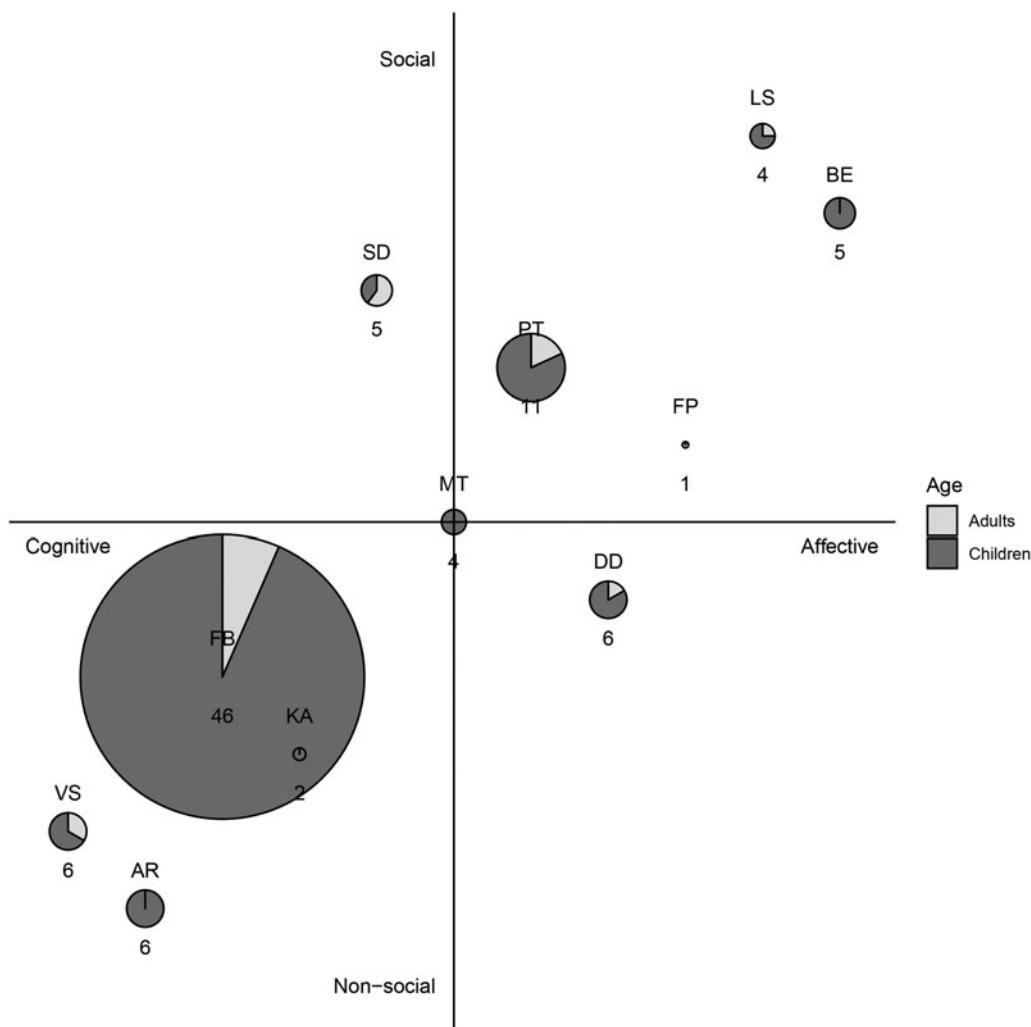
#### Language of the tasks

We examined the languages used in administration. Six studies were unclear on which language the task was delivered in, but a classification was inferred based on the bilinguals' L1 (first

language) and L2 (second language)<sup>5</sup> and indications in the text. Approximately 30% ( $n = 16$ ) of the publications tested bilinguals in both languages. Another 8% ( $n = 4$ ) offered the participant a choice of language they preferred to be tested. The remainder only assessed bilinguals using one of their languages, either L1 ( $n = 8$ , 15%) or L2 ( $n = 17$ , 32%). In five cases (9%), bilinguals' L1 and L2 varied across the sample. An example would be Tiv et al. (2021) which tested all participants in Québec in English, where English could be either L1 or L2.

#### Testing and response modalities

The testing and response modalities were coded as verbal or non-verbal abilities. Verbal tasks are defined as tasks requiring participants to process speech or text. Nonverbal tasks do not require participants to rely on speech or text, such as watching a soundless cartoon or looking at pictures, though verbal instructions can be given. For responses, a verbal response encompasses speaking or writing, and a nonverbal response is pointing to pictures. Thirty-eight studies used verbal tasks (72%), four used non-verbal (8%), and eleven (21%) studies combined both verbal and non-verbal tasks. An example of mixing verbal and nonverbal formats would be administering the unexpected transfer task (Wimmer & Perner, 1983) with both a verbal and a nonverbal trial, as documented in Sudo and Matsui (2021). The verbal tasks were generally variations on the false-belief tasks, such as unexpected contents, unexpected location, and change-in-location. We found that 29 studies included some form of object or picture to supplement the verbal task, primarily for children, though some of the nonverbal tasks do include verbal instructions or stories. For instance, Ryskin et al. (2014) used complex map and grid tasks with pictures and nonverbal cues for the participant to navigate. However, the task requires verbal navigational directions given by the experimenter, like the Director Task. To be fully classified as a non-verbal task, these tasks usually involve visuospatial orientation where, while verbal instructions were given in the trial, the task itself is entirely visual with pictures (e.g., Gorrell, 1987; Greenberg et al., 2013).



**Figure 2.** Visual Representation of Types of ToM Tasks by Frequency of Occurrence in the Database.  
 Note. Size of the pie charts is proportional to the frequency counts of studies in the database. AR = Appearance Reality, BE = Belief Emotion, DD = Diverse Desire and Belief, FB = False Belief, FP = Faux Pas, KA = Knowledge Access, LS = Language Switching, MT = Multiple, PT = Perspective Taking, SD = Scene Description, VS = Visuospatial.

Regarding response format, we found that 39 studies (74%) required participants to respond verbally, and 7 (13%) required mixed verbal and nonverbal responses. Twenty-eight of these studies used simple verbal responses such as a simple statement. These are generally associated with false-belief tasks. Other verbal responses are more complex such as open-ended questions and story retelling. To illustrate, Grøver (2019) had children retell a story from the perspective of two different characters in a picture book. Han and Lee (2013) also used a story retell task, however, theirs involved taking away pictures from a set and asking participants to retell the story with missing information. Nonverbal responses were primarily studies that included pictures in a multiple-choice answer format. We also noted a few studies that used eye-tracking assessing gaze when engaging in perspective-taking (e.g., Rubio-Fernández & Glucksberg, 2012; Ryskin et al., 2014).

*Cognitive and affective ToM*

We further looked at whether the studies distinguished between cognitive ToM and affective ToM. Recall that cognitive ToM focuses on thoughts and beliefs, while affective ToM involves

emotions and feelings. Note that this classification and its results should be carefully interpreted because the distinction we made between cognitive and affective tasks is based on whether the task involved emotions or not. Around 77% (n = 41) of studies focused only on cognitive ToM. Certain studies included both cognitive and affective ToM in one task (n = 6, 11%) or separate tasks (n = 4, 8%). Only two studies used tasks that assessed only affective ToM (in other words, affective ToM was their primary construct).

*Bilingual characterization*

The studies in the database were examined to see how the researchers operationalize monolingualism and bilingualism. Researchers defined the groups on one of the following methods: home or school languages (n = 7, 13%), language proficiency (n = 4, 8%), self-reports (n = 5, 8%), or through questionnaires used and assessed by the authors (n = 16, 30%). However, most studies used a mixture of different methods (n = 16, 30%) to classify bilinguals. For instance, Chan (2005) first identified monolingual and bilingual children through language use in daycares. The

classification was confirmed by parents' reports, which was further supported through a language proficiency assessment to identify balanced bilinguals.

Notably, the criteria set by each researcher all have considerable variability from study to study. For instance, both Yow and Markman (2015) and Nguyen and Astington (2014) established a strict numeric criterion for language exposure for their developing samples. Yow and Markman defined bilinguals as individuals with a minimum of 30% exposure to one of their languages WEEKLY, while Nguyen and Astington specified 30% exposure to each of the two languages IN GENERAL. Kovács (2009) defined bilinguals by using daily exposure to both languages with no specific cut-off point. Further, 7 studies (13%) did not specify how they differentiated monolinguals and bilinguals (see Kobayashi et al., 2006, 2007, 2008). One study determined bilinguals archaically with cultural/ethnic identity: either Vietnamese–American or Spanish–American (Gorrell et al., 1982). It is also important to note that monolingual groups may also show diverse language exposure. For instance, Tarighat and Krott (2021) reported monolingual adults with L2 age of acquisition, daily use, and years of L2 exposure.

### Language proficiency

We found that 87% ( $n = 46$ ) of the studies reported assessing language proficiency, of which 5 used self-reports. Of these 46 studies, the majority ( $n = 27$ , 51%) examined the proficiency of both L1 and L2 for the bilingual sample. The remaining ( $n = 11$ , 21%) examined proficiency in either L1 ( $n = 2$ ) or L2 ( $n = 9$ ). The language of assessment was not reported in three studies. The 38 studies that used formal assessment varied in the measurement they used. Predominantly, studies used the Peabody Pictures Vocabulary Test and its translations like the Spanish Test de Vocabulario en Imágenes Peabody. Other measurements were the MacArthur Communicative Development Inventory, Test of Early Language Development-3 and Receptive One Word Picture Vocabulary Test.

### Language history, exposure, and context

Most studies did not adopt developed questionnaires such as aLSBQ (Anderson et al., 2018; Luk & Bialystok, 2013) and LEAP-Q (Marian et al., 2007) when assessing participants' language history. We made a qualitative assessment of the presence of three features of bilingual experience: language history, language exposure, and language context. Approximately half of the papers reported language history ( $n = 28$ , 53%), and a slightly smaller proportion ( $n = 25$ , 47%) reported the amount of language exposure. We saw the opposite in dissertations, where a detailed report of the participants and the questionnaire used to assess language background were often included.

Regarding language context, we examined it in two different methods, separated by age groups. First, we found that 74% of the studies ( $n = 39$ ) reported contexts where language interactions occurred. When organized by age, 33 of the 42 studies on children (79%) reported some information on language context. Among the adult studies, we found 6 of the 10 studies detailed the contexts. We then assessed the 39 studies to determine what kinds of contexts were discussed. We identified that the studies discuss the school ( $n = 4$ ) and home language contexts ( $n = 5$ ), with a majority discussing both ( $n = 23$ , 21 studies on children). We also noted that several publications detailed the language context at a more societal level ( $n = 18$ , 34% of all 53 papers). Five of the 10 adult studies discussed the social context in detail. Tiv et al.

(2021), for instance, contextualized the bilinguals, detailing the official languages and the neighbourhood diglossia unique to those living in Montréal, Québec.

## Discussion

This scoping review examined the literature on multilinguals' Theory of Mind (ToM). We specifically looked at the ToM tasks administered to the bilingual population for children and adults. The majority (79%) of the studies looked at children instead of adults. Characterization of language experiences for both children and adolescents was present but limited. Across the 53 publications we reviewed, there is a heavy reliance on false-belief tasks to measure ToM. The tasks in the literature also heavily focused on a non-social cognitive dimension of ToM. We discuss major findings and their implications on research on bilingualism and ToM: (1) the challenges of studying ToM in bilingual adults; (2) why the characterization of bilinguals can be helpful; (3) clarification between types of ToM; and (4) concerns regarding ToM task administration modalities. We conclude this discussion by outlining potential future directions for investigating ToM in multilingual adults.

### A gap on adult ToM

We found that the studies in the database concentrated on ToM development in children (79%). This finding is understandable as studies have shown that ToM develops around the age of four (Wellman et al., 2001). Researchers gravitate towards developing samples, a group that is acquiring both languages and ToM. However, there is limited literature examining ToM in adult bilinguals. Specifically, only eleven studies examined adults (individuals aged 18 or over), and only one looked at the aging population (see Cox et al., 2016). ToM does not "finish" developing upon reaching four years old. Adults also often fail to take on the perspective of someone else, as shown using cognitive ToM tasks (e.g., Apperly et al., 2010; Bernstein et al., 2011; Birch & Bloom, 2004, 2007; Epley et al., 2004; Keysar et al., 2003). Some researchers have attributed this failure to the CURSE OF KNOWLEDGE, a tendency to bias one's own knowledge (e.g., Birch & Bloom, 2004). Others associated the findings with executive functioning, where adults struggle with maintaining the perspective of another individual when performing the tasks (e.g., Apperly et al., 2010). With only eleven studies on bilingual adults, we believe future research can address this existing gap and develop age-appropriate ToM tasks for adults.

Bernstein et al. (2017) outlined two difficulties with studying ToM across the lifespan. The first limitation is due to the ceiling effect of existing ToM tasks that were designed for children. Bernstein et al. posit that the tasks like Sally-Anne assess ToM as an "all-or-none" phenomenon – as a categorical variable. As such, adults may successfully pass a ToM task and show little variability for quantitative analyses. Tasks designed for adults require an adaptation to detect differences in an age-appropriate setting. To address this concern, some tasks, such as the Sandbox task (Sommerville et al., 2013), used a spatial layout, thus allowing ToM to be measured as a continuous variable. In our database, researchers generally used more complex tasks for adults that evaluate higher-order ToM (e.g., "where would Sally think Anne thinks where Sally thinks the ball is?"). Several studies required more complex instructions, such as the map and grid task in Ryskin et al. (2014). We saw one instance of Faux Pas



(Cox et al., 2016), which required a more advanced understanding of social norms (though this task has also been used with children). These measurements are more age-appropriate for adults as children would not have the linguistic and social capabilities to pass these tasks successfully.

Another issue lies in the fact that ToM is not an isolated ability; ToM is connected to factors like executive functioning and, most importantly, language (Bernstein et al., 2017). Therefore, assessing ToM specifically in adults can be challenging. Taking bilingual adults' diverse lived experiences into consideration, ToM studies can benefit from considering the variability in other factors (i.e., executive functioning, memory, language history, and language abilities) that arise across the lifespan. For example, Rubio-Fernández and Geurts (2013) showed that certain verbal false-belief tasks depend on attention: through the conscious tracking of another's perspective, children can pass the ToM task successfully. In this review, we noted that about half (45%) of the studies evaluated some form of executive functioning. Furthermore, we argue that the finding is still limited to the cognitive domains of ToM. Understanding adult ToM beyond just cognitive ToM can help understand how social perspective-taking and shifting occur in practical scenarios. Similarly, bilingual adults' ToM requires consideration of other factors such as metalinguistic abilities and language history. Therefore, we conceptualize a third challenge in studying adults' ToM intersecting with the characterization of bilinguals: language history. Adult bilinguals present a uniquely rich linguistic experience. Some may have learned their second language as an adult after moving to a new country, while others may have been bilingual all their life but never used their heritage language outside their family. It is still unclear how the heterogeneity of bilingual experiences modulates ToM. Many aspects of bilingualism are not reported in the studies investigating ToM, considering that we found seven studies overlooked assessing language proficiency, and 25 (47%) studies did not collect or detail any information on language history. One way to resolve this difficulty is a deeper characterization of the people behind the study (see next section). We echo Apperly et al. (2009) that studies on adult ToM can inform us about ToM development.

### *Sociolinguistic contexts and multilingual characterization*

When measuring ToM, aspects such as sociolinguistic context become even more relevant since sociolinguistic awareness is one of the three prominent explanations for bilinguals' improved ToM (Díaz, 2021; Yu et al., 2021). If the argument is that bilinguals are more sensitive to social cues due to their lived experience, ToM should be evaluated in relevant social contexts. Yet, when we assess ToM, the participants are isolated in the situational vacuum of a false-belief task. These tasks do not represent the naturalistic contexts in which bilinguals exercise their ToM, especially considering that ToM is a social phenomenon. We consider the concept of SOCIAL ToM to be essential in a bilingual context. Our results indicated a heavy reliance on said cognitive tasks. A dependence on cognitive tasks does not fully encapsulate bilinguals' ToM, given that ToM inherently relies on social processes. Rarely do false-belief tasks or appearance-reality tasks involve interacting with another active agent. Taking another person's perspective requires engagement with another individual rather than reading story vignettes or answering multiple-choice questions.

Note, however, that several studies did use ToM tasks that were grounded in social settings. Cheung et al. (2010) explicitly

examined sociolinguistic awareness, the only study in the database to do so. They described their bilingual child participants as second language learners in the sociolinguistic context of Hong Kong, where language switching between Cantonese and English can occur frequently. For their task, experimenters simply greeted the children in the opposite language: they greeted them in English if the child said "hello" in Cantonese, and vice versa. The authors categorized the children into two groups: bilinguals, children who used English frequently due to an English-only staff at school; and second language learners, children that only used English in English lessons. They found that sociolinguistic awareness – operationalized as the number of attempts in the language switching prompt – predicted ToM, more so for bilingual children than second language learners. This design was a simple task to examine whether children would switch their language, situated in a context familiar to these children. Although other studies did not overtly state they were assessing sociolinguistic factors, they used tasks in the context where bilinguals may switch their languages. Kovács (2009) used a modified ToM task that mimicked a language-switching scenario by having the children infer false beliefs based on the language of the interaction. Tare and Gelman (2010) did a similar task during free play that required the children to be aware of the experimenter's language and switch appropriately during play. These different tasks incorporated ToM in social interaction, allowing bilingual participants to use their language and ToM in a more natural context.

Additionally, a social ToM task should coexist with a considerate characterization of the bilingual participants' social experiences to help understand and justify how these experiences relate to ToM development. Studies that include a detailed characterization of social contexts help capture the multilingual experience, as one's ability to speak may vary depending on the language and social contexts (Navarro-Torres et al., 2021). Context is vital to consider, especially under the model of the Adaptive Control Hypothesis (Green & Abutalebi, 2013), which addresses how different interactional contexts frame bilingual language control processes. Grosjean's complementarity principle (1997, 2010) indicates equivalently that the context in which we use our language varies. In the *Systems Framework of Bilingualism*, Titone and Tiv (2022) support the need to detail social levels of language use, from individual backgrounds to interpersonal interactions. Therefore, the context in which bilinguals use their languages would result in different perspectives and levels of control processes and, by inference, different amounts of engagement with ToM. These discussions also highlight the importance of detailing bilingual factors such as language acquisition history, proficiency, and contexts, variables that are standards in bilingual research as studies move from a categorical characterization of bilinguals into a more multidimensional approach (Byers-Heinlein et al., 2019; Surrain & Luk, 2019).

Certain studies in the database did characterize their bilingual population. For example, Sudo and Matsui (2021) detailed the language environment of Brazilian children living and attending school in Japan. The authors outlined the dual language situation, the school setting, and the socioeconomic challenges that may affect these children's language learning, use, and academics. Through this characterization, we can understand the sociolinguistic context and how the results affect this group of bilingual individuals. However, our review found that only 34% of the studies detailed the sociolinguistic contexts in which bilingualism occurred, similar to the percentage reported by Surrain and Luk (2019) with studies not limited to ToM. Several recent papers in the database

did not characterize the language experience in detail (e.g., Gasiorok et al., 2022; Kim et al., 2021; Stegall-Rodriguez et al., 2021). The heterogeneity in qualifiers of bilinguals makes interpreting the results from group comparisons difficult.

Recently, a study used network science to characterize bilinguals (Navarro et al., 2022<sup>6</sup>). This method examined relationships, experiences, and social networks. Although lengthy, network science shows promise as a holistic and meticulous methodology in detailing bilinguals' language history and interactions. The measurement used by Navarro et al. was the Director Task, a cognitive ToM task focused on perspective-taking (also arguably attention; see Rubio-Fernández, 2017). While the task involves a more social aspect (due to the inclusion of a "director" in the task), the task itself is still limited by its laboratory design. Another way to evaluate sociolinguistic influence is addressed in the recent study by Surrain and Luk (2023). The authors captured local multilingualism using US Census data coupled with survey respondents' IP addresses and a detailed parent home language questionnaire. The presence of local multilingualism played a significant role in predicting the perceived value of bilingualism. These are some ways to help characterize bilingual participants and their sociolinguistic contexts.

### Clarifying the types of ToM

Most significantly, we found 96 ToM tasks used across the 53 publications. Studies would often modify various tasks to suit their research purposes. We highlighted how the unexpected contents tasks had seven separate citations. This finding exemplifies the concern outlined by Warnell and Redcay (2019), where ToM tasks designed to measure the same construct fail to correlate with each other; participants who perform well on one task may not perform as well on another. With so many variations among just 53 studies, it is arguable whether these tasks are measuring the same type of ToM.

For this review, we addressed the diversity of tasks by categorizing them into 11 types of ToM tasks. Through this categorization, we can see a heavy reliance on false-belief as a measurement of ToM (46 cases). This scope narrows the examination of ToM when instead, we should broaden and understand the various facets of ToM development among bilinguals. For example, there were only four cases of tasks focused on language switching, a more socially-oriented ToM. Language-switching tasks require participants to take the perspective in a communicative scenario, which is exceedingly different from a visuospatial task where the bilingual participants are rotating objects. We are not indicating that all studies should include as many ToM measurements as possible. Although we saw four cases where a singular task included several types of ToM (e.g., Banasik & Podsiadło, 2016), these tasks have been mainly some form of standardized testing such as the Theory of Mind Inventory-2 (ToMI-2; Hutchins et al., 2012) used by Kim et al. (2021). These measurements are also limited in scope and might be normed to linguistically homogeneous samples or may not apply to adult populations.

In addition to evaluating ToM situated in social settings, the distinction between cognitive and affective ToM can clarify the aspect of ToM that bilinguals develop differently. Studies in other topics or disciplines, such as adults with schizophrenia (e.g., Yeh et al., 2021) and neuroscience (e.g., Healey & Grossman, 2018), make this differentiation between these two types of ToM. While it is not a novel concept (see studies dating

as early as Kurdek & Rodgon, 1975), we found that 23% of studies assessed affective ToM. The mental engagement of cognitive ToM differs from the emotional perspective of affective ToM. Under the model proposed by Shamay-Tsoory et al. (2010), cognitive ToM is a prerequisite to affective ToM and serves as an implication in other aspects, such as empathy. In this review, two studies attempted to make this direct differentiation (Han & Lee, 2013; Rodriguez, 2000). Han and Lee (2013) observed between-group differences in the affective task but not in the cognitive one. The remaining studies used an affective task to evaluate emotions and feelings (e.g., Mante-Estacio & Bernardo, 2015), but they ultimately did not clarify the difference between cognitive and affective ToM. Future research is needed to clarify this difference, perhaps with neuroimaging studies that show the distinct neural activation between cognitive and affective ToM (e.g., Shamay-Tsoory et al., 2009). A selective design in subsequent studies can expand on Han and Lee's results (2013) by distinguishing the specific type of ToM that bilinguals develop differently.

### ToM task administration and response modalities

ToM tasks can be confounded with language abilities (for both bilinguals and monolinguals, see discussion in Bloom & German, 2000) as language proficiency is necessary to abstractly represent the ToM scenarios (de Villiers & Pyers, 2002; Pyers & Senghas, 2009). For example, several studies used and cited tasks from Wellman and Liu (2004). Consider an example of the EXPLICIT FALSE BELIEF task. Children are given a story about Scott:

Here's Scott. Scott wants to find his mittens. His mittens might be in his backpack or they might be in the closet. *Really*, Scott's mittens are in his backpack. But Scott *thinks* his mittens are in the closet. So, where will Scott look for his mittens? (Wellman & Liu, 2004, p. 539; emphasis by original authors).

Bilingual participants, whose English may not be their first language, must be able to pick up on the words "really," "but," and "thinks" to pass successfully. Therefore, this task becomes a language assessment above and beyond one of ToM. In this scoping review, we noted that 17 studies (32%) assessed bilingual participants in their second language (L2). Note that due to the limited characterization of the bilingual participants, L2 could be the preferred language of these participants (once again reinforcing the importance of characterization). However, under the assumption that L2 could be the less frequent or less proficient language, the participants can struggle with the ToM tasks. Gordon (2016) directly addressed proficiency and found that bilingual proficiency in BOTH languages predicted higher performance on various ToM tasks. Therefore, language proficiency can play a vital role in the performance of ToM tasks. To address this concern, studies can evaluate the proficiency of the bilingual participants to not only statistically control for proficiency but also contextualize their findings. However, of the 46 studies that did evaluate proficiency, 14 assessed only one language. Furthermore, seven studies did not evaluate proficiency and tested bilingual participants in their L2.

We did find that around 30% of the studies assessed bilinguals' ToM in both languages. However, there are some challenges with this approach. For instance, Goetz (2003) tested both languages, but monolinguals were tested twice to match the bilingual

participants. Therefore, despite a noticeable bilingual difference at time 1, monolinguals performed better at time 2 (due to possible practice effect), and thus the between-group differences were gone when including time 2 scores. Nguyen and Astington (2014) attempted the same method and indicated in their supplemental material that they doubled the scores for bilinguals to match the monolinguals. These manipulations can affect the interpretation of the results.

Assessing in both languages is not always possible. Translating tasks into another language can be problematic. For instance, consider a group of balanced bilinguals with L1 Chinese and L2 English in Lee et al. (1999). When tasks used verbs like 以为, *yi wei*, which directly imply false belief, resulted in a better performance than tasks using another verb like 想, *xiang*, which is closely translated as “to think”. Translating tasks would be inappropriate without accounting for these linguistic differences. Therefore, a possible solution is to let the bilinguals choose the language they prefer to be tested in, such as in Sperling (1990) and Díaz and Farrar (2018a, 2018b). While this proposal does not solve the proficiency and language processing difficulties, it could be a way to move forward in testing as it takes into account individual preferences. Realistically, not all studies can benefit from a bilingual group using the same two languages (consider a Canadian sample where 3+ languages are common).

Another area of concern we observed was the matching of modalities. To illustrate, consider the three studies that used questionnaires instead of tasks: Barber et al. (2021) used a subtest from NEPSY-II, Tarighat and Krott (2021) used the Interpersonal Reactivity Index (IRI; Davis, 1980), and Kim et al. (2021) used the Theory of Mind Inventory-2. Barber et al. (2021) assessed reading abilities as their study focused on language comprehension. Using a reading task such as the NEPSY-II was suitable for their study because the appropriate language ability was considered and tested. However, being bilingual does not imply that they are biliterate (being able to read in both languages). Therefore, when a ToM task is selected, studies should match the language assessment with the modality of that task, similar to how Barber et al. (2021) addressed their study. Tasks like the Interpersonal Reactivity Index rely on participants reading and understanding the written prompts. If bilingual participants are classified as bilinguals based on a verbal assessment, it is insufficient to assume they would pass a reading task in a second language. It is essential to consider whether the task modality suits the bilingual sample involved.

Finally, Greenberg et al. (2013) argued that due to bilinguals' lower language abilities in their second developing language when compared to monolinguals, non-verbal tasks might be more appropriate. We found that 28% of the studies we reviewed used a nonverbal component (either alone or in combination with a verbal task). The purely nonverbal tasks are cognitive and rely on visual-based perspective-taking. Therefore, we believe it is still debatable whether purely nonverbal tasks are suitable for assessing ToM. Hollebrandse et al. (2014) showed that monolingual 7-year-old children could successfully pass the verbal ToM task but not the low-verbal version of the task, further showing support that language plays a role in ToM performance. Statistical control for language abilities could sufficiently address the issue (see Gallant et al., 2020; Warnell & Redcay, 2019).

### Future directions

ToM tasks are grounded in the fact that we take on the perspective of another social being. Whether it is to consider their visual

perception or their emotions, we disengage our egocentrism and step into the shoes of someone else. Therefore, to fully study bilinguals' ToM, there must be naturalistic interaction in the sociolinguistic context that bilinguals inhabit. There have been ethnographic observational studies (e.g., Kremer-Sadlik, 2004) and naturalistic conversations (Alkire et al., 2022) in ToM literature. For example, the idea of conversational ToM (cToM) conceptualized by Alkire et al. (2022) could be a possible measurement as it scores conversations based on references to the other person's mental state. The authors prompted a conversation, asking, “why don't we get to know each other?”. While Alkire et al. only evaluated children, the task can also be developmentally appropriate for adults. This task, combined with the detailed characterization of social networks, could be an example of future direction given that the interaction between the individual and their environment exemplifies what it means to be bilingual (Luk & Grundy, 2022).

Every study has its objective and ways to operationalize its relevant constructs. We echo Byers-Heinlein et al. (2019) in that “Researchers will need to select and adapt questions to their own research questions, study protocol, and populations” (p. 6). It depends on the theoretical framework and the field in which researchers will approach these topics. Cognitive scientists may feel inclined to continue examining ToM through a quantitative lens. Neuroscientists may focus on identifying the neural networks of cognitive and affective ToM. Social scientists may consider a more qualitative approach and extract themes from naturalistic conversations and adopt a mixed-methods approach. Here, we aim to provide some considerations when researching bilinguals' ToM across disciplines. We conceptualize these points as a set of suggestions rather than enforced rules.

1. A careful characterization of the bilingual participants is beneficial to understanding ToM, particularly in sociolinguistic contexts in which the participants use multiple languages socially.
2. Studies can adopt diverse methodologies. ToM can be examined beyond false beliefs and encompasses other socio-cognitive representations as well.
3. If the objective of a study is to examine factors associated with bilinguals' ToM, then measurements of executive functioning, metalinguistic awareness, and language abilities would benefit from theoretical relevance.
4. Following guidelines provided by Quesque and Rossetti (2020), the selection of ToM task(s) should consider the level of cognitive processes as they relate to ToM.

### Limitations

This scoping review took a broad evaluation of the literature on bilinguals' ToM. As such, it is primarily a qualitative and descriptive analysis of the existing papers. A systematic review or meta-analysis with more stringent criteria could focus on synthesizing the actual results from these studies (see Yu et al., 2021). Alternatively, upon completing this review, we acknowledge that a separate set of criteria may be necessary. For instance, studies on children with hearing disabilities can inform ToM research (de Villiers, 2005), but this review opted not to include these studies due to diverse findings in samples with developmental disorders.

Another limitation is that we have not discussed the complexity of more advanced ToM (AToM; e.g., Osterhaus et al., 2016), which is the understanding of higher-order ToM. These types

of ToM may be worth exploring as these levels are much more complex as they require an individual to consider multiple and more distant perspectives. The suggestion regarding sociolinguistic tasks and characterization may not be relevant in the context of AToM. It is uncertain how naturalistic third and fourth-order ToM can be, but it remains a consideration for future reviews.

While previous research argues that ToM develops similarly across cultures (Liu et al., 2008; Wellman et al., 2001), recent evidence suggests otherwise (e.g., Wu et al., 2013). Some studies in the database considered culture as part of one's language background (e.g., Navarro & Conway, 2021) – however, cultural considerations remain as another factor to incorporate. Similarly, another unique result was the correlation between personality (conscientiousness specifically) with perspective-taking (see Lorge & Katsos, 2019). No other studies considered personality in the database, and this variable could influence social ToM as some individuals may seek out more interactions and perspectives. In the context of bilinguals, personality (and the fact they speak more languages) could result in a similar social context rich with interactions and perspectives, which warrants further examination.

## Conclusion

Both understanding the development of ToM and the consequence of bilingualism can benefit from reevaluating and expanding the methodology and study design. In this scoping review, we illustrated several considerations and trends in the current literature and identified several gaps for future research. The previous research has helped inform us of the cognitive consequence of the bilingual experience, including those observed in ToM in the existing literature. Investigation ToM from a social or affective perspective may be a natural next step by incorporating sociolinguistic contexts of bilingualism and examining its consequences. This direction is particularly meaningful for adults with qualitatively different social experiences with multiple languages.

**Acknowledgements.** This research was funded by Research Grants from the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPIN-2020-05052) to the third author and the Graduate Student Stipend from the Centre for Research on Brain, Language & Music (CRBLM) to the first author. CRBLM is funded by the Government of Quebec via the Fonds de Recherche Nature et Technologies and Société et Culture.

**Supplementary Material.** For supplementary material accompanying this paper, visit <http://dx.doi.org/10.1017/S1366728923000585>

## Notes

1 Appearance-reality involves using objects that resembles (appears) as one thing, but is actually something else (reality).

2 Faux-pas tasks asks participants to read a story vignette involve characters performing something that is against social norms.

3 These counts were approximate because not all studies described their procedure in detail, others have made significant modifications to warrant their own version.

4 Some authors did not provide a reference to the unexpected-contents task.

5 It is important to note here that while tasks were administered in L1 and L2, it is only an indication of their age of acquisition by the authors. The results here do not indicate whether the bilinguals are more proficient in their L1 or L2.

6 Note this study was not included in the database due as our literature search date was on January 24<sup>th</sup>, 2022.

\* Note for References section: publications with asterisk are part of the scoping review database

## References

- Alkire, D., McNaughton, K. A., Yarger, H. A., Shariq, D., & Redcay, E. (2022). Theory of mind in naturalistic conversations between autistic and typically developing children and adolescents. *Autism*, 13623613221103700. <https://doi.org/10.1177/13623613221103699>
- Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, 50(1), 250–263. <https://doi.org/10.3758/s13428-017-0867-9>
- Anderson, J. A. E., Hawrylewicz, K., & Bialystok, E. (2020). Who is bilingual? Snapshots across the lifespan. *Bilingualism: Language and Cognition*, 23(5), 929–937. <https://doi.org/10.1017/S1366728918000950>
- Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology*, 45(1), 190–201. <https://doi.org/10.1037/a0014098>
- Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, 63(6), 1201–1217. <https://doi.org/10.1080/17470210903281582>
- Astington, J. W., & Baird, J. A. (2005). *Why language matters for Theory of Mind*. Oxford University Press. <http://ebookcentral.proquest.com/lib/mcgill/detail.action?docID=281341>
- Banasik, N., & Podsiadlo, K. (2016). Comprehension of ironic utterances by bilingual children. *Psychology of Language and Communication*, 20(3), 316–335. Scopus. <https://doi.org/10.1515/plc-2016-0019>\*
- Barber, A. T., Vizcaya-Jofré, F., & Klauda, S. L. (2021). The importance of theory of mind in oral and reading comprehension in emergent bilingual students. *Pensamiento Educativo*, 58(2). Scopus. <https://doi.org/10.7764/PEL.58.2.2021.3>\*
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37–46. [https://doi.org/10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8)
- Batres, K. M. (2014). Bilingual cognitive control and perspective-monitoring in dialogue. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 75(1-B(E)), No-Specified.\*
- Bensalah, L., Caillies, S., & Anduze, M. (2016). Links among cognitive empathy, theory of mind, and affective perspective taking by young children. *The Journal of Genetic Psychology*, 177(1), 17–31. <https://doi.org/10.1080/00221325.2015.1106438>
- Berguno, G., & Bowler, D. M. (2004). Communicative interactions, knowledge of a second language, and theory of mind in young children. *The Journal of Genetic Psychology*, 165(3), 293–309.\*
- 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>
- Bernstein, D. M., Coolin, A., Fischer, A. L., Thornton, W. L., & Sommerville, J. A. (2017). False-belief reasoning from 3 to 92 years of age. *PLOS ONE*, 12(9), e0185345. <https://doi.org/10.1371/journal.pone.0185345>
- Bialystok, E., & Senman, L. (2004). Executive processes in appearance-reality tasks: The role of inhibition of attention and symbolic representation. *Child Development*, 75(2), 562–579. <https://doi.org/10.1111/j.1467-8624.2004.00693.x>\*
- Bigelow, F. J., Clark, G. M., Lum, J. A. G., & Enticott, P. G. (2021). The mediating effect of language on the development of cognitive and affective theory of mind. *Journal of Experimental Child Psychology*, 209, 105158. <https://doi.org/10.1016/j.jecp.2021.105158>
- 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. *Psychological Science*, 18(5), 382–386. <https://doi.org/10.1111/j.1467-9280.2007.01909.x>
- Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, 77(1), B25–B31. [https://doi.org/10.1016/S0010-0277\(00\)00096-2](https://doi.org/10.1016/S0010-0277(00)00096-2)

- Buac, M., & Kaushanskaya, M. (2020). Predictors of Theory of Mind performance in bilingual and monolingual children. *International Journal of Bilingualism*, 24(2), 339–359. Scopus. <https://doi.org/10.1177/1367006919826866>\*
- Byers-Heinlein, K., Esposito, A. G., Winsler, A., Marian, V., Castro, D. C., & Luk, G. (2019). The case for measuring and reporting bilingualism in developmental research. *Collabra: Psychology*, 5(1), 37. <https://doi.org/10.1525/collabra.233>
- Cassetta, B. D., Pexman, P. M., & Goghari, V. M. (2018). Cognitive and affective theory of mind and relations with executive functioning in middle childhood. *Merrill-Palmer Quarterly*, 64(4), 514. <https://doi.org/10.13110/merrpalmquar1982.64.4.0514>
- Chan, K. T. (2005). Chinese-English bilinguals' theory-of-mind development. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 65(10-A), 3688.\*
- Cheung, H., Mak, W. Y., Luo, X., & Xiao, W. (2010). Sociolinguistic awareness and false belief in young Cantonese learners of English. *Journal of Experimental Child Psychology*, 107(2), 188–194. <https://doi.org/10.1016/j.jecp.2010.05.001>\*
- Cox, S. R., Bak, T. H., Allerhand, M., Redmond, P., Starr, J. M., Deary, I. J., & MacPherson, S. E. (2016). Bilingualism, social cognition and executive functions: A tale of chickens and eggs. *Neuropsychologia*, 91, 299–306. <https://doi.org/10.1016/j.neuropsychologia.2016.08.029>\*
- Dahlgren, S., Almen, H., & Dahlgren Sandberg, A. (2017). Theory of mind and executive functions in young bilingual children. *The Journal of Genetic Psychology*, 178(5), 303–307. <https://doi.org/10.1080/00221325.2017.1361376>\*
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85.
- de Villiers, P. A. (2005). The role of language in Theory-of-Mind development: What deaf children tell us. In J. W. Astington & J. A. Baird (Eds.), *Why Language Matters for Theory of Mind* (pp. 266–297). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195159912.003.0013>
- de Villiers, J. G., & de Villiers, P. A. (2014). The role of language in theory of mind development. *Topics in Language Disorders*, 34(4), 313–328. <https://doi.org/10.1097/TLD.0000000000000037>
- de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, 17(1), 1037–1060. [https://doi.org/10.1016/S0885-2014\(02\)00073-4](https://doi.org/10.1016/S0885-2014(02)00073-4)
- Díaz, V. (2021). Minds in action: Evidence that linguistic diversity helps children build a theory of mind. *Bilingualism: Language and Cognition*, 25(1), 1–11. <https://doi.org/10.1017/S1366728921000109>
- Díaz, V., & Farrar, M. J. (2018a). Do bilingual and monolingual preschoolers acquire false belief understanding similarly? The role of executive functioning and language. *First Language*, 38(4), 382–398. <https://doi.org/10.1177/0142723717752741>\*
- Díaz, V., & Farrar, M. J. (2018b). The missing explanation of the false-belief advantage in bilingual children: A longitudinal study. *Developmental Science*, 21(4), 1–13.\*
- Dicataldo, R., & Roch, M. (2020). Are the effects of variation in quantity of daily bilingual exposure and socioeconomic status on language and cognitive abilities independent in preschool children? *International Journal of Environmental Research and Public Health*, 17(12), 1–23. <https://doi.org/10.3390/ijerph17124570>\*
- Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13(2), 331–338. <https://doi.org/10.1111/j.1467-7687.2009.00888.x>
- Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40(6), 760–768. <https://doi.org/10.1016/j.jesp.2004.02.002>
- Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science*, 26(7), 1090–1097. <https://doi.org/10.1177/0956797615574699>\*
- Farhadian, M., Abdullah, R., Mansor, M., Redzuan, M., Gazanizadand, N., & Kumari, V. (2010). Theory of Mind in bilingual and monolingual preschool children. *Journal of Psychology*, 1(1), 39–46. <https://doi.org/10.1080/09764224.2010.11885444>\*
- Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, 79(1), 126–138. <https://doi.org/10.1111/j.1467-8624.2007.01115.x>
- Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive Psychology* 15(1), 95–120. [https://doi.org/10.1016/0010-0285\(83\)90005-1](https://doi.org/10.1016/0010-0285(83)90005-1)
- Frank, T. I. (2000). The use of word-learning principles in young monolingual and bilingual children. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 61(1-B), 562.\*
- Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, 50(4), 531–534. <https://doi.org/10.1016/j.neuron.2006.05.001>
- Gallant, C. M. M., Lavis, L., & Mahy, C. E. V. (2020). Developing an understanding of others' emotional states: Relations among affective theory of mind and empathy measures in early childhood. *British Journal of Developmental Psychology*, 38(2), 151–166. <https://doi.org/10.1111/bjdp.12322>
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493–501. [https://doi.org/10.1016/S1364-6613\(98\)01262-5](https://doi.org/10.1016/S1364-6613(98)01262-5)
- Gasiorek, J., Dragojevic, M., & Vincze, L. (2022). Perspective-taking and language competence as predictors of language accommodation by adolescents from monolingual and bilingual households. *International Journal of Bilingual Education and Bilingualism*, 25(1), 148–155. Scopus. <https://doi.org/10.1080/13670050.2019.1641467>\*
- Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition*, 6(1), 1–15. <https://doi.org/10.1017/S1366728903001007>\*
- Gordon, K. R. (2016). High proficiency across two languages is related to better mental state reasoning for bilingual children. *Journal of Child Language*, 43(2), 407–424. <https://doi.org/10.1017/S0305000915000276>\*
- Gorrell, J. (1987). Spatial role-taking ability among bilingual and monolingual kindergarten children. *Journal of Psycholinguistic Research*, 16(2), 91–99. <https://doi.org/10.1007/BF01071997>\*
- Gorrell, J. J., Bregman, N. J., McAllister, H. A., & Lipscomb, T. J. (1982). A comparison of spatial role-taking in monolingual and bilingual children. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 140(1), 3–10. <https://doi.org/10.1080/00221325.1982.10534170>\*
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5), 515–530. <https://doi.org/10.1080/20445911.2013.796377>
- Greenberg, A., Bellana, B., & Bialystok, E. (2013). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive Development*, 28(1), 41–50. <https://doi.org/10.1016/j.cogdev.2012.10.002>\*
- Grosjean, F. (1997). The bilingual individual. *Interpreting: International Journal of Research and Practice in Interpreting*, 2, 163–187.
- Grosjean, F. (2010). *Bilingual*. Harvard University Press.
- Grosjean, F. (2013). Bilingualism: A short introduction. In *The psycholinguistics of bilingualism* (pp. 5–25). John Wiley & Sons.
- Grøver, V. (2019). Young bilinguals' extended discourse skills: The role of perspective taking. In *Learning through Language: Towards an Educationally Informed Theory of Language Learning* (pp. 179–191). Scopus. <https://doi.org/10.1017/9781316718537.017>\*
- Han, S., & Lee, K. (2013). Cognitive and affective perspective-taking ability of young bilinguals in South Korea. *Child Studies in Diverse Contexts*, 3(1), 69–80. <https://doi.org/10.5723/csdc.2013.3.1.069>\*
- Healey, M. L., & Grossman, M. (2018). Cognitive and affective perspective-taking: Evidence for shared and dissociable anatomical substrates. *Frontiers in Neurology*, 9, 491. <https://doi.org/10.3389/fneur.2018.00491>
- Hollebrandse, B., van Hout, A., & Hendriks, P. (2014). Children's first and second-order false-belief reasoning in a verbal and a low-verbal task. *Synthese*, 191(3), 321–333. <https://doi.org/10.1007/s11229-012-0169-9>
- Hsin, L., & Snow, C. (2017). Social perspective taking: A benefit of bilingualism in academic writing. *Reading and Writing: An Interdisciplinary Journal*, 30(6), 1193–1214. <https://doi.org/10.1007/s11145-016-9718-9>\*
- Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2012). Psychometric evaluation of the Theory of Mind Inventory (ToMI): A study of typically

- developing children and children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 42(3), 327–341. <https://doi.org/10.1007/s10803-011-1244-7>
- Jean-Louis, B. (1999). Does childhood bilingualism facilitate the development of theory of mind? *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 60(5-B), 2386.\*
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on Theory of Mind use in adults. *Cognition*, 89(1), 25–41. [https://doi.org/10.1016/S0010-0277\(03\)00064-7](https://doi.org/10.1016/S0010-0277(03)00064-7)
- Kim, Y.-S. G., Wolters, A., Mercado, J., & Quinn, J. (2021). Crosslinguistic transfer of higher order cognitive skills and their roles in writing for English-Spanish dual language learners. *Journal of Educational Psychology*, No-Specified. <https://doi.org/10.1037/edu0000516>\*
- Kobayashi, C., Glover, G. H., & Temple, E. (2006). Cultural and linguistic influence on neural bases of “Theory of Mind”: An fMRI study with Japanese bilinguals. *Brain and Language*, 98(2), 210–220.\*
- Kobayashi, C., Glover, G. H., & Temple, E. (2007). Cultural and linguistic effects on neural bases of “Theory of Mind” in American and Japanese children. *Brain Research*, 1164(0045503, b5), 95–107.\*
- Kobayashi, C., Glover, G. H., & Temple, E. (2008). Switching language switches mind: Linguistic effects on developmental neural bases of “Theory of Mind”. *Social Cognitive and Affective Neuroscience*, 3(1), 62–70. <https://doi.org/10.1093/scan/nsm039>\*
- Kovács, A. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, 12(1), 48–54. <https://doi.org/10.1111/j.1467-7687.2008.00742.x>
- Kremer-Sadlik, T. (2004). How children with autism and asperger syndrome respond to questions: A ‘naturalistic’ theory of mind task. *Discourse Studies*, 6(2), 185–206. <https://doi.org/10.1177/1461445604041767>
- Kurdek, L. A., & Rodgon, M. M. (1975). Perceptual, cognitive, and affective perspective taking in kindergarten through sixth-grade children. *Developmental Psychology*, 11(5), 643. <https://doi.org/10.1037/0012-1649.11.5.643>
- Kyuchukov, H., & De Villiers, J. (2009). Theory of Mind and evidentiality in Romani-Bulgarian bilingual children. *Psychology of Language and Communication*, 13(2), 21–34. Scopus. <https://doi.org/10.2478/v10057-009-0007-4>\*
- Lee, J. Y. S., & Imuta, K. (2021). Lying and Theory of Mind: A meta-analysis. *Child Development*, 92(2), 536–553. <https://doi.org/10.1111/cdev.13535>
- Lee, K., Olson, D. R., & Torrance, N. (1999). Chinese children’s understanding of false beliefs: The role of language. *Journal of Child Language*, 26(1), 1–21. <https://doi.org/10.1017/S0305000998003626>
- Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, 44(2), 523. <https://doi.org/10.1037/0012-1649.44.2.523>
- Lohmann, H., & Tomasello, M. (2004). The role of language in the development of false belief understanding: A training study. *Child Development*, 74(4), 1130–1144. <https://doi.org/10.1111/1467-8624.00597>
- Lorge, I., & Katsos, N. (2019). Listener-adapted speech: Bilinguals adapt in a more sensitive way. *Linguistic Approaches to Bilingualism*, 9(3), 376–397. Scopus. <https://doi.org/10.1075/lab.16054>\*
- Luk, G. (2022). Justice and equity for whom? Reframing research on the “bilingual (dis)advantage.” *Applied Psycholinguistics*, 1–15. <https://doi.org/10.1017/S0142716422000339>
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25(5), 605–621. <https://doi.org/10.1080/20445911.2013.795574>
- Luk, G., & Grundy, J. G. (2022). The importance of recognizing social contexts in research on bilingualism. *Bilingualism: Language and Cognition*, 1–3. <https://doi.org/10.1017/S1366728922000177>
- Mante-Estacio, M. J., & Bernardo, A. B. I. (2015). Illusory transparency in bilinguals: Does language of text affect bilingual readers’ perspective taking in reading? *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, 34(4), 744–752. <https://doi.org/10.1007/s12144-014-9286-6>\*
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). *The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals*. 28.
- Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, 78(2), 622–646. <https://doi.org/10.1111/j.1467-8624.2007.01018.x>
- Navarro-Torres, C. A., Beatty-Martínez, A. L., Kroll, J. F., & Green, D. W. (2021). Research on bilingualism as discovery science. *Brain and Language*, 222, 105014. <https://doi.org/10.1016/j.bandl.2021.105014>
- Navarro, E. (2021). Theory of mind measurements and mechanisms: An investigation of construct validity and cognitive processes in theory of mind tasks. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 82(11-B), No-Specified.\*
- Navarro, E., & Conway, A. R. (2021). Adult bilinguals outperform monolinguals in theory of mind. *Quarterly Journal of Experimental Psychology* (2006), 74(11), 1841–1851. <https://doi.org/10.1177/17470218211009159>\*
- Navarro, E., DeLuca, V., & Rossi, E. (2022). It takes a village: Using network science to identify the effect of individual differences in bilingual experience for theory of mind. *Brain Sciences*, 12(4), 487. <https://doi.org/10.3390/brainsci12040487>
- Nguyen, T.-K., & Astington, J. W. (2014). Reassessing the bilingual advantage in theory of mind and its cognitive underpinnings. *Bilingualism: Language and Cognition*, 17(2), 396–409. <https://doi.org/10.1017/S1366728913000394>\*
- Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of advanced Theory-of-Mind tasks. *Child Development*, 87(6), 1971–1991. <https://doi.org/10.1111/cdev.12566>
- Pearson, D. K. (2013). *Effect of Language Background on Metalinguistic Awareness and Theory of Mind*. 236.\*
- Pelletier, J. (2006). Relations among theory of mind, metacognitive language, reading skills and story comprehension in L1 and L2 learners. In *Theory of mind and language in developmental contexts*. (pp. 77–92). Springer Science+Business Media, Inc.\*
- Poletti, M., Enrici, I., & Adenzato, M. (2012). Cognitive and affective Theory of Mind in neurodegenerative diseases: Neuropsychological, neuroanatomical and neurochemical levels. *Neuroscience & Biobehavioral Reviews*, 36(9), 2147–2164. <https://doi.org/10.1016/j.neubiorev.2012.07.004>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515–526. <https://doi.org/10.1017/S0140525X00076512>
- Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding. *Psychological Science*, 20(7), 805–812. <https://doi.org/10.1111/j.1467-9280.2009.02377.x>
- Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, 15(2), 384–396. <https://doi.org/10.1177/1745691619896607>
- Raisa, Z. H., Karuppali, S., Bhat, J. S., & Bajaj, G. (2019). Thinking about what he thinks of what I think: Assessing higher theory of mind abilities in Indian bilingual children between 3.0 and 8.11 years of age. *Indian Journal of Psychiatry*, 61(2), 167–176. [https://doi.org/10.4103/psychiatry.IndianJPsychiatry\\_115\\_18](https://doi.org/10.4103/psychiatry.IndianJPsychiatry_115_18)\*
- Rodriguez, M. D. (2000). Language proficiency, perspective taking and the empathic understanding of Hispanic bilingual children. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 60(9-B), 4922.\*
- Rothman, J., Bayram, F., DeLuca, V., Pisa, G. D., Duñabeitia, J. A., Gharibi, K., Hao, J., Kolb, N., Kubota, M., Kupisch, T., Laméris, T., Luque, A., Osch, B., van, Soares, S. M. P., Prystauka, Y., Tat, D., Tomić, A., Voits, T., & Wulff, S. (2022). Monolingual comparative normativity in bilingualism research is out of “control”: Arguments and alternatives. *Applied Psycholinguistics*, 1–14. <https://doi.org/10.1017/S0142716422000315>
- Rubio-Fernández, P. (2017). Why are bilinguals better than monolinguals at false-belief tasks? *Psychonomic Bulletin & Review*, 24(3), 987–998. <https://doi.org/10.3758/s13423-016-1143-1>
- 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., & Glucksberg, S. (2012). Reasoning about other people’s beliefs: Bilinguals have an advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(1), 211–217. <https://doi.org/10.1037/a0025162>\*
- Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T. (2014). Visuospatial perspective-taking in conversation and the role of

- bilingual experience. *Journal of Memory and Language*, 74, 46–76. <https://doi.org/10.1016/j.jml.2014.04.003>\*
- Schick, B., De Villiers, P., De Villiers, J., & Hoffmeister, R. (2007). Language and Theory of Mind: A study of deaf children. *Child Development*, 78(2), 376–396. <https://doi.org/10.1111/j.1467-8624.2007.01004.x>
- Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, 45(13), 3054–3067. <https://doi.org/10.1016/j.neuropsychologia.2007.05.021>
- Shamay-Tsoory, S. G., Tomer, R., Berger, B. D., Goldsher, D., & Aharon-Peretz, J. (2005). Impaired “affective theory of Mind” is associated with right ventromedial prefrontal damage. *Cognitive and Behavioral Neurology*, 18(1), 55–67. <https://doi.org/10.1097/01.wnn.0000152228.90129.99>
- Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* 132(3), 617–627. <https://doi.org/10.1093/brain/awn279>
- Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex*, 46(5), 668–677. <https://doi.org/10.1016/j.cortex.2009.04.008>
- Singh, L., Moh, Y., Ding, X., Lee, K., & Quinn, P. C. (2021). Cognitive flexibility and parental education differentially predict implicit and explicit racial biases in bilingual children. *Journal of Experimental Child Psychology*, 204(2985128r, i2n), 105059. <https://doi.org/10.1016/j.jecp.2020.105059>\*
- Smetana, J. G., Jambon, M., Conry-Murray, C., & Sturge-Apple, M. L. (2012). Reciprocal associations between young children’s developing moral judgments and theory of mind. *Developmental Psychology*, 48(4), 1144–1155. <https://doi.org/10.1037/a0025891>
- Smogorzewska, J., Szumski, G., Bosacki, S., Grygiel, P., & Karwowski, M. (2022). School engagement, sensitivity to criticism and academic achievement in children: The predictive role of theory of mind. *Learning and Individual Differences*, 93, 102111. <https://doi.org/10.1016/j.lindif.2021.102111>
- 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>
- Sperling, M. T. (1990). Social-cognitive development of bilingual and monolingual children. *Dissertation Abstracts International*, 50(11-B), 5349.\*
- Stegall-Rodriguez, S. E., Weimer, A. A., & Rice Warnell, K. (2021). Alternative perspectives: Relations between belief reasoning and ambiguous figure perception in bilingual children. *Infant and Child Development*, No-Specified. <https://doi.org/10.1002/icd.2258>\*
- Stich, S., & Ravenscroft, I. (1994). What is folk psychology? *Cognition*, 50(1), 447–468. [https://doi.org/10.1016/0010-0277\(94\)90040-X](https://doi.org/10.1016/0010-0277(94)90040-X)
- Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to Theory of Mind. *Journal of Cognitive Neuroscience* 10(5), 640–656. <https://doi.org/10.1162/089892998562942>
- Sudo, M., & Matsui, T. (2021). School readiness in language-minority dual language learners in Japan: Language, executive function, and theory of mind. *The Journal of Genetic Psychology*, 182(6), 375–390. <https://doi.org/10.1080/00221325.2021.1930994>\*
- Surrain, S., & Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*, 22(2), 401–415. <https://doi.org/10.1017/S1366728917000682>
- Surrain, S., & Luk, G. (2023). The perceived value of bilingualism among U.S. parents: The role of language experience and local multilingualism. *Translational Issues in Psychological Science*, 1–12. <https://doi.org/10.1037/tps0000352>
- Tare, M., & Gelman, S. A. (2010). Can you say it another way? Cognitive factors in bilingual children’s pragmatic language skills. *Journal of Cognition and Development: Official Journal of the Cognitive Development Society*, 11(2), 137–158.\*
- Tarighat, S., & Krott, A. (2021). Bilingualism enhances reported perspective taking in men, but not in women. *Frontiers in Psychology*, 12(101550902), 679524. <https://doi.org/10.3389/fpsyg.2021.679524>\*
- Titone, D. A., & Tiv, M. (2022). Rethinking multilingual experience through a systems framework of bilingualism. *Bilingualism: Language and Cognition*, 1–16. <https://doi.org/10.1017/S1366728921001127>
- Tiv, M., O’Regan, E., & Titone, D. (2021). In a bilingual state of mind: Investigating the continuous relationship between bilingual language experience and mentalizing. *Bilingualism: Language and Cognition*, 918–931. <https://doi.org/10.1017/S1366728921000225>\*
- Tricco, A. C., Lillie, E., Zarin, W., O’Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, 191, 103997. <https://doi.org/10.1016/j.cognition.2019.06.009>
- Watson, A. C., Painter, K. M., & Bornstein, M. H. (2001). Longitudinal relations between 2-year-olds’ language and 4-year-olds’ Theory of Mind. *Journal of Cognition and Development*, 2(4), 449–457. [https://doi.org/10.1207/S15327647JCD0204\\_5](https://doi.org/10.1207/S15327647JCD0204_5)
- Weimer, A. A., & Gasquoine, P. G. (2016). Belief reasoning and emotion understanding in balanced bilingual and language-dominant Mexican American young children. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 177(2), 33–43. <https://doi.org/10.1080/00221325.2016.1138793>\*
- Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development*, 75(2), 523–541. <https://doi.org/10.1111/j.1467-8624.2004.00691.x>
- 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>
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition*, 13(1), 103–128. [https://doi.org/10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5)
- Wimmer, M. C., Marx, C., Stirk, S., & Hancock, P. J. B. (2021). Bilinguals’ inhibitory control and attentional processes in a visual perceptual task. *Psychological Research*, 85(4), 1439–1448. <https://doi.org/10.1007/s00426-020-01333-0>\*
- Wu, S., Barr, D. J., Gann, T. M., & Keysar, B. (2013). How culture influences perspective taking: Differences in correction, not integration. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00822>
- Yeh, Y.-C., Lin, C.-Y., Li, P.-C., Hung, C.-F., Cheng, C.-H., Kuo, M.-H., & Chen, K.-L. (2021). A systematic review of the current measures of Theory of Mind in adults with schizophrenia. *International Journal of Environmental Research and Public Health*, 18(13), 7172. <https://doi.org/10.3390/ijerph18137172>
- Yow, W. Q., & Markman, E. M. (2015). A bilingual advantage in how children integrate multiple cues to understand a speaker’s referential intent. *Bilingualism: Language and Cognition*, 18(3), 391–399. <https://doi.org/10.1017/S1366728914000133>\*
- Yu, C.-L., Kovelman, I., & Wellman, H. M. (2021). How bilingualism informs Theory of Mind development. *Child Development Perspectives*, 15(3), 154–159. <https://doi.org/10.1111/cdep.12412>