



# Systematic Review and Inventory of Theory of Mind Measures for Young Children

Cindy Beaudoin<sup>1,2</sup>, Élizabel Leblanc<sup>1</sup>, Charlotte Gagner<sup>1,2</sup> and Miriam H. Beauchamp<sup>1,2\*</sup>

<sup>1</sup> Department of Psychology, University of Montreal, Montreal, QC, Canada, <sup>2</sup> Sainte-Justine Hospital Research Center, Montreal, QC, Canada

Theory of mind (TOM), the ability to infer mental states to self and others, has been a pervasive research theme across many disciplines including developmental, educational, neuro-, and social psychology, social neuroscience and speech therapy. TOM abilities have been consistently linked to markers of social adaptation and have been shown to be affected in a broad range of clinical conditions. Despite the wealth and breadth of research dedicated to TOM, identifying appropriate assessment tools for young children remains challenging. This systematic review presents an inventory of TOM measures for children aged 0–5 years and provides details on their content and characteristics. Electronic databases (1983–2019) and 9 test publisher catalogs were systematically reviewed. In total, 220 measures, identified within 830 studies, were found to assess the understanding of seven categories of mental states and social situations: emotions, desires, intentions, percepts, knowledge, beliefs and mentalistic understanding of non-literal communication, and pertained to 39 types of TOM sub-abilities. Information on the measures' mode of presentation, number of items, scoring options, and target populations were extracted, and psychometric details are listed in summary tables. The results of the systematic review are summarized in a visual framework “Abilities in Theory of Mind Space” (ATOMS) which provides a new taxonomy of TOM sub-domains. This review highlights the remarkable variety of measures that have been created to assess TOM, but also the numerous methodological and psychometric challenges associated with developing and choosing appropriate measures, including issues related to the limited range of sub-abilities targeted, lack of standardization across studies and paucity of psychometric information provided.

**Keywords:** theory of mind, systematic review, childhood, psychometrics, assessment, preschool

## OPEN ACCESS

### Edited by:

Ilaria Grazzani,  
University of Milano Bicocca, Italy

### Reviewed by:

Diane Poulin-Dubois,  
Concordia University, Canada  
Manuel Sprung,  
Karl Landsteiner University of Health  
Sciences, Austria

### \*Correspondence:

Miriam H. Beauchamp  
miriam.beauchamp@umontreal.ca

### Specialty section:

This article was submitted to  
Developmental Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 13 June 2019

**Accepted:** 09 December 2019

**Published:** 15 January 2020

### Citation:

Beaudoin C, Leblanc É, Gagner C and  
Beauchamp MH (2020) Systematic  
Review and Inventory of Theory of  
Mind Measures for Young Children.  
*Front. Psychol.* 10:2905.  
doi: 10.3389/fpsyg.2019.02905

## INTRODUCTION

Consolidating appropriate social skills is an essential part of typical development, as it allows individuals to establish and maintain satisfying social relationships and promotes community adaptation across the lifespan (Cacioppo, 2002). The emergence of social skills is a complex developmental process involving the maturation of a broad range of underlying cognitive functions, referred to as “social cognition” (Beauchamp and Anderson, 2010). Among these, Theory of Mind (TOM) has been a central focus of developmental and social psychology, as well as speech therapy (Byom and Turkstra, 2012) since Premack first coined the term TOM in the 1970s, referring to

the ability to impute mental states to self and others, including desires, knowledge, beliefs, and intentions, in order to predict behavior (Premack and Woodruff, 1978). In order to display flexible and explicit TOM, it was acknowledged that children must have the capacity to construct different abstract representations of reality, and to navigate between them to distinguish their mental states from those of others using various cues, therefore acting as “theorists” (Wimmer and Perner, 1983). This field has since been one of the most studied in developmental cognitive science (Sabbagh and Paulus, 2018). More recently, TOM and other social cognitive constructs have also attracted attention within the field of social neuroscience, which has generated a large body of consensual literature regarding the brain networks underlying TOM (Gallagher and Frith, 2003; Frith and Frith, 2006; Blakemore, 2008; Bellerose et al., 2011; Bird and Viding, 2014).

Children who have good TOM generally display markers of social adaptation, such as better communication skills, better quality social relationships, increased peer popularity and higher academic achievement (Binnie, 2005; Fink et al., 2015; Slaughter, 2015; Slaughter et al., 2015; Imuta et al., 2016). Conversely, poorer TOM has been identified in a number of conditions and contexts characterized by altered social functioning, such as autism spectrum disorders (Yirmiya et al., 1998; Shaked and Yirmiya, 2004; Senju, 2012; Chung et al., 2014; Kimhi, 2014; Leekam, 2016), language impairment (Stanzione and Schick, 2014), attention-deficit/hyperactivity disorder (Bora and Pantelis, 2016), Tourette’s syndrome (Eddy and Cavanna, 2013), childhood maltreatment (Luke and Banerjee, 2013; Benarous et al., 2015), conduct disorders (Anastassiou-Hadjicharalambous and Warden, 2008; Poletti and Adenzalo, 2013), anorexia nervosa (Bora and Köse, 2016), schizophrenia (Brune, 2005; Sprong et al., 2007; Bora et al., 2009; Cermolacce et al., 2011; Biedermann et al., 2012; Chung et al., 2014; Martin et al., 2014; Song et al., 2015; Healey et al., 2016), traumatic brain injury (Snodgrass and Knott, 2006; Walz et al., 2010; Dennis et al., 2012; McDonald, 2013; Bellerose et al., 2017), epilepsy (Bora and Meletti, 2016; Stewart et al., 2016), neurofibromatosis (Payne et al., 2016), and Fragile X syndrome (Turkstra et al., 2014).

Efforts to understand the role of TOM in normative development and in clinical conditions are ongoing. Furthering this knowledge relies on the use of validated, developmentally appropriate assessment tools, especially given that social cognition is now included in the assessment recommendations of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V; American Psychiatric Association, 2013). Although a surfeit of measures have been developed to test TOM (particularly in the field of cognitive science), identifying the best measure for particular clinical or research needs is not an easy enterprise. Evaluating TOM presents many challenges, some of which are related to the numerous and varied definitions and conceptualisations of TOM that have been proposed (Premack and Woodruff, 1978; Wimmer and Perner, 1983; Leslie, 1987; Tager-Flusberg and Sullivan, 2000; Abu-Akel and Shamay-Tsoory, 2011; Dennis et al., 2013; Bird and Viding, 2014; Westby, 2014; Asakura and Inui, 2016; Happé et al., 2017), the changeable manifestations of TOM at different developmental

stages (Wellman et al., 2011; Carlson et al., 2013; Slaughter, 2015), and the psychometric limitations associated with some measures (Mayes et al., 1996; Brune, 2001; Hutchins et al., 2008a; Carlson et al., 2013; Hiller et al., 2014).

## Defining Theory of Mind and Distinguishing It From Other Social Constructs

TOM is a complex construct encompassing a range of abilities, which are variably targeted as a function of the measurement tool chosen (German and Cohen, 2012). Each definition or theory provides slightly different conceptions regarding the specificity of TOM and what behavioral manifestations it reflects (Premack and Woodruff, 1978; Wimmer and Perner, 1983; Leslie, 1987; Tager-Flusberg and Sullivan, 2000; Abu-Akel and Shamay-Tsoory, 2011; Dennis et al., 2013; Bird and Viding, 2014; Westby, 2014; Asakura and Inui, 2016; Happé et al., 2017). Nonetheless, it is generally accepted that TOM represents a set of cognitive skills that enable reasoning about cognitive (e.g., beliefs) or affective (e.g., emotions) mental states.

In this review, the Self to Other Model of Empathy (SOME; Bird and Viding, 2014) is used as a framework to define TOM and set the inclusion and exclusion criteria for the literature search. The SOME is a comprehensive model based on empirical data from clinical and neuroimaging studies (Bird and Viding, 2014). It depicts how social cognitive constructs, such as TOM, come together to determine empathic behavior rather than focusing solely on internal TOM processes. Importantly, SOME distinguishes TOM from empathy: TOM is defined as a person’s cognitive representation of self and other’s mental states, whereas empathy is defined as an emotional contagion caused by exposure to another’s emotion, while being conscious that this emotional state is experienced by the other (Bird and Viding, 2014). In the model, TOM is also differentiated from the “affective cue classification system,” a lower perceptual system responsible for processing and categorizing stimuli signaling affective states, such as facial emotions and tones of voice. The SOME model further posits that TOM is distinct from a “situation understanding system” responsible for processing situational cues and deducing or associating estimated emotional states of others based upon situational cues (e.g., people dressed in black at a cemetery = funeral = sadness) (Bird and Viding, 2014). The model is therefore useful for setting boundaries between TOM and other closely related social cognitive constructs, and was used in the current review to distinguish central TOM measures from those more distally related to TOM.

In addition to using a clear definition of TOM to identify and document relevant assessment tools, the construct of TOM should be distinguished from other abilities that, though they may build or rely on TOM, are better represented by other social cognitive functions. For example, many overt prosocial and self-promoting behaviors rely on TOM, but can be more directly assessed through targeted measures, such as those that document cooperation, adherence to social norms, lies and manipulative interpersonal tactics (Baurain and Nader-Grosbois, 2013; Slaughter, 2015). The way in which TOM is used in everyday social interactions also depends on other discrete

factors, such as temperament, life experiences, integration of social values and executive functioning (Beauchamp and Anderson, 2010; Slaughter, 2015; Vera-Estay et al., 2015). As a result, in order to identify assessment measures that specifically target TOM, it is also critical to choose those that elicit TOM specifically, rather than those that evaluate more complex social cognitive skills, such as moral reasoning (Vera-Estay et al., 2015) and strategic social decision making (Steinmann et al., 2014), for example.

There are developmental considerations that should also be taken into account to constrain our search to the most unambiguous forms of TOM. There is ongoing debate around the definition of TOM with regards to which emerging social skills in infancy are considered direct, early manifestations of TOM, and which are distinct cognitive precursors allowing TOM to arise (Carlson et al., 2013). While the question of the first measurable manifestations of TOM remains to be answered theoretically and empirically, current literature and most authors suggest that early social skills, such as imitation, gaze following, pointing, and joint attention, may reflect, at most, more automatic, implicit manifestations of awareness of mental states (Carlson et al., 2013). These skills are thus thought to act as precursors of later-developing TOM skills that reflect an explicit, coherent, flexible and conceptual understanding of mental states (Carlson et al., 2013), and that constitute the topic of the current review. In sum, this review constrains TOM so as to distinguish it from empathy, classification of affective and situational cues, early non-explicit cognitive representations of mental states, such as joint attention and imitation, and more complex social abilities, such as cooperation or manipulation tactics.

## The Developmental Trajectory of TOM and Associated Measurement Tools

Taking into account the diverse definitions and conceptions of TOM, it is not surprising that a broad variety of paradigms and measures have been developed to study the construct. Despite the range of mental states a child must learn to interpret (e.g., emotions, knowledge, intents, beliefs, desires), there appears to be an over-representation of measures directed specifically at assessing one particular type of mental state: false beliefs (Hedger and Fabricius, 2011; Hiller et al., 2014). The false belief paradigm was initially proposed by Wimmer and Perner (1983) and has since been adapted and applied to a range of contexts (Wellman et al., 2001). Typically, children are presented with a short scenario depicting a contradiction between reality and a character's belief. For example, in the change of location paradigm referred to as the Sally and Ann task (Baron-Cohen et al., 1985), two dolls, Sally and Ann, are presented to a child. Sally places her marble in a basket, and then leaves the scene. Ann takes the marble out of the basket and puts it in a box. When Sally comes back, the child is asked where she would search for the marble. To succeed in this task, children have to answer "in the basket," despite the fact that they know that the marble is really in the box. This type of scenario enables experimenters to determine

a child's ability to understand that a person's mental state is not a simple reflection of reality, and suggests that the child is able to elaborate a theory about another person's mental content, a "theory of mind".

Children typically complete false belief paradigms successfully somewhere between 3 and 5 years of age (Wellman et al., 2001), an observation which has long been linked to the assumption that this is the period during which TOM develops. However, the use of a broader variety of measures and methods has subsequently shown that TOM follows a more extended and nuanced developmental trajectory (Wellman et al., 2011). In particular, the emergence of implicit, non-verbal and simplified measures designed to be used in very young, pre-verbal infants, suggested that some TOM abilities may already be present in infancy, a conclusion that could not be reached using standard measures because of the extraneous factors inherent to the tests (Slaughter, 2015). For example, these studies used implicit methods, such as observation of imitation behaviors, violation-of-expectation paradigms and eye gaze tracking to show that children demonstrate some knowledge of the intentions of others around 12–18 months of age (Kristen et al., 2011), can appreciate others' desires around 18 months of age (Repacholi and Gopnik, 1997; Poulin-Dubois et al., 2007), and show some comprehension of false beliefs as early as 15 months of age (Onishi and Baillargeon, 2005; Southgate et al., 2007; Senju, 2012). The interpretation of these results has been the subject of much debate: whereas some claim that implicit tasks are valid methods to measure TOM (Carruthers, 2013; Powell et al., 2018), others suggest that they lack reliability and validity data to support their use (Dörrenberg et al., 2018; Kulke et al., 2018). This debate has been fueled by failed attempts to replicate studies using implicit measures of false-belief understanding, leading to a "replication crisis" (Sabbagh and Paulus, 2018). The issue of the reliability and validity of these tasks is intertwined with that of the nature of what is measured using implicit methods to test "theory of mind," contributing to the debate regarding the conception and development of TOM and its first measurable manifestations (Heyes, 2014; Scott and Baillargeon, 2017; Sabbagh and Paulus, 2018). Conversely, the use of a variety of more complex explicit TOM tasks has suggested that TOM continues to develop after the age of 5 years. For example, children improve on their ability to understand second order false belief tasks (i.e., "Ann thinks that Sally thinks the marble is in the basket") between 5 and 6 years of age, and develop an increasingly mature appreciation of sarcasm, *faux-pas* (social gaffes) and white lies throughout adolescence (Miller, 2009). Neuroimaging studies also depict longitudinal changes in patterns of cerebral activation during a variety of TOM tasks, and suggest protracted development well through adolescence and into adulthood (Blakemore, 2008, 2012). Together, these findings highlight that TOM cannot be seen as a unitary construct and must be appreciated in light of its ongoing development. They also support the importance of relying on diverse TOM measures that are reliable, valid and sensitive to developmental changes in order to adequately document a complex and rapidly changing cognitive ability.

## Psychometric Challenges Associated With TOM Measures

Despite significant advances in our understanding of both normative and altered TOM (Wellman et al., 2001; Gallagher and Frith, 2003; Vuadens, 2005; Poletti and Adenzalo, 2013; Kimhi, 2014; Imuta et al., 2016), it is still difficult to draw robust conclusions about its role in typical development and clinical conditions. Such challenges may be the result of the methodological weaknesses associated with measures used to assess TOM (Hiller et al., 2014; Henry et al., 2016). Indeed, the psychometric standards of TOM measures have been qualified as unsystematic, suboptimal, and immature (Mayes et al., 1996; Brune, 2001; Hutchins et al., 2008a; Carlson et al., 2013; Hiller et al., 2014). The methodological weaknesses of TOM assessment include reliance on measures with one or two tests items only (Cutting and Dunn, 1999; Garner et al., 2005), over-representation of false belief understanding as the sole measure of TOM (Wellman and Liu, 2004; Carlson et al., 2013; Hiller et al., 2014), and the fact that few TOM measures have empirically validated psychometric properties (Hutchins et al., 2008a; Hiller et al., 2014; Ziatabar Ahmadi et al., 2015).

## Existing Sources of Information on TOM Measures

To our knowledge, no systematic review has been conducted to document the characteristics of existing TOM measures for young children. Non-systematic reviews have been published on TOM measures that are widely used in clinical populations (Sprung, 2010), in adulthood (Henry et al., 2015), and in middle childhood and adolescence (Hayward and Homer, 2017). These reviews highlight the relevance of a number of TOM measures for understanding social functioning in clinical conditions and typical development and provide interesting insights in the ways to use them, but they are not systematic and do not cover tools destined for infants, toddlers and preschoolers. Ziatabar Ahmadi et al. (2015) conducted a systematic review of TOM measures for preschoolers, but constrained the scope to articles presenting the development and validation of comprehensive measures composed of multiple TOM tasks. Therefore, their review excludes single task measures (e.g., single false belief tasks) that constitute the majority of measures used in TOM research (Hiller et al., 2014). In addition, the review conducted by Ziatabar Ahmadi et al. (2015) is limited to studies that specifically aim to validate the psychometric properties of TOM measures, thus excluding other types of empirical studies (e.g., longitudinal, outcome or prediction papers).

The primary objective of this study was to systematically record an inventory of existing measures that assess TOM in children under the age of 6 years of age (0–5 years). This age range was chosen because the period between 3 and 5 years is widely recognized as a sensitive period for TOM development (Wellman et al., 2001). The range was extended down to infancy because there is no actual consensus regarding the age at which the first manifestations of TOM appear (Carlson et al., 2013). This inventory will assist researchers and clinicians in choosing

measures that best fit their needs and will identify possible gaps or limits inherent to existing measures.

## METHODS

A systematic review of the literature was conducted. Empirical studies referring to TOM measures used with young children were reviewed using a search protocol based on The Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (PRISMA; Moher et al., 2015). Eligibility criteria were pre-determined both at the level of study selection and identification of TOM measure (see **Table 1** for the list of eligibility criteria and associated exclusion criteria).

## Sources of Information and Search Strategy

The search strategy was created in collaboration with a psychology librarian. The following electronic databases were searched: Ovid PsycINFO, Health and Psychosocial Instruments, MEDLINE(R) In-Process and Other Non-Indexed Citations and MEDLINE(R). The dates of coverage were from 1983 to October 2019. The start date (1983) was chosen because of seminal work published in that year (Wimmer and Perner, 1983).

The following key search terms, pertaining to children (1), measures (2), and TOM (3) were used, in combination, and restrained to “all journals”:

1. (child\* or schoolchild\* or toddler\* or preschool\* or infan\*).mp [mp = title, abstract, heading word, table of contents, key concepts, original title, tests, and measures]
2. (psychometric\* or validation or questionnaire\* or scale\* or inventor\* or instrument\* or measure\* or tool or assess\* or evaluation\*).mp [mp = title, abstract, heading word, table of contents, key concepts, original title, tests, and measures]
3. (theory of mind or false belief\* or perspective taking\* or social attribution\* or belief attribution\* or desires reasoning).mp [mp = title, abstract, heading word, table of contents, key concepts, original title, tests, and measures]

In addition to the standard electronic search databases, the catalogs of the following English or French publishers of testing materials were manually reviewed: Pearson Assessment Canada, Psychological Assessment Resources, Institut de Recherches Psychologiques, Western Psychological Services, Hogrefe, Les Éditions du Centre de Psychologie Appliquée, Eurotests Editions, PsychTest, Schuhfried. Whenever the age range of participants could not be extracted directly from an article, the corresponding author was contacted to obtain the information. Moreover, whenever the cited source of an assessment tool was not retrieved using the search strategy, it was manually searched and included as a record to be screened alongside others in the selection process, even though it was published before 1983.

## Selection Process

Search results were imported to an Endnote X7 database. Screening was performed in two phases. In phase 1, all search results were screened for the eligibility criteria based only on

**TABLE 1** | Eligibility and exclusion criteria for the systematic review.

Eligibility Criteria	Exclusion criteria
The document is accessible, in its full version, at the time of the search	<b>Missing document:</b> The document's reference is incomplete and does not allow identification of the full text or full text could not be found <b>Unpublished:</b> The document is not published or in press, or the in-press content is not accessible
The document is written in English or French. A list of possibly relevant titles in other languages is provided as <b>Appendix I</b>	<b>Other language:</b> The document is written in another language than English or French
The document is from a peer-reviewed journal or published by a test publisher/editor	<b>Non-reviewed or non-commercialized:</b> The document is not published in a peer reviewed journal (e.g., books, book chapters, conference proceedings, are excluded), nor is it commercialized
The document reports the results of an empirical study providing original data	<b>Not an empirical study:</b> The records do not report a study providing new and original data (e.g., theoretical articles, literature reviews, meta-analyses, letters, editorials, etc.)
The measure is used with human subjects	<b>Non-human subjects:</b> The measure was not administered to humans (e.g., animal studies)
The measure is administered to young children (<6 years of age). Studies with participants 6 years of age or over are included, provided the sample is also composed of children under 6 years. The sample may be composed of adults, as long as the measure aims to evaluate TOM in a child under 6 years of age (e.g., parental report in the form of a questionnaire)	<b>Older participants:</b> The measure is not administered to a child under 6 years of age
The measure provides a score or a classification. Subjective (i.e., questionnaires) or objective (i.e., direct testing, observational coding systems) are included	<b>Lack of score:</b> The measure does not provide a score or classification reflecting an individual's TOM (e.g., research paradigms used to solicit TOM during neuroimaging, but that do not score the participant's TOM abilities)
The measure can be used to assess TOM in normative or clinical conditions (physical, psychological or neurological)	<b>Narrow utility:</b> The measure is useful only in the case of a specific condition (e.g., blindness)
The measure aims to evaluate TOM as defined in the introduction, that is the ability to create a cognitive representation of self and other's mental states (SOME; Bird and Viding, 2014)	<b>No TOM or diverging TOM definition:</b> The measure does not assess TOM or does not assess it in a way that is consistent with the chosen theoretical framework (Bird and Viding, 2014). Measures that assess a more complex social behavior or ability (e.g., moral reasoning), a precursor social cognitive skill (e.g., joint attention), or another social cognitive construct (e.g., empathy, affective cues classification) are excluded

the content of the title and abstract, by two of the authors. Two decisions were possible at this stage: exclusion based on an eligibility criterion or inclusion for phase 2. In phase 2, the full texts of all remaining search results were screened for

eligibility criteria by three of the authors. Two decisions were possible at this stage: exclusion based on an eligibility criterion or inclusion in the systematic review. For each phase, the first 15% of search results were screened independently by all reviewers in order to obtain an inter-rater agreement in terms of inclusion or exclusion of the search result. The inter-rater agreement was 89.9% at phase 1 and 93.9% at phase 2. During the entire process, any discrepancies or difficulties in the identification of inclusion/exclusion criteria were resolved by discussion with the other reviewers and authors if needed.

## Content Analysis and Data Extraction

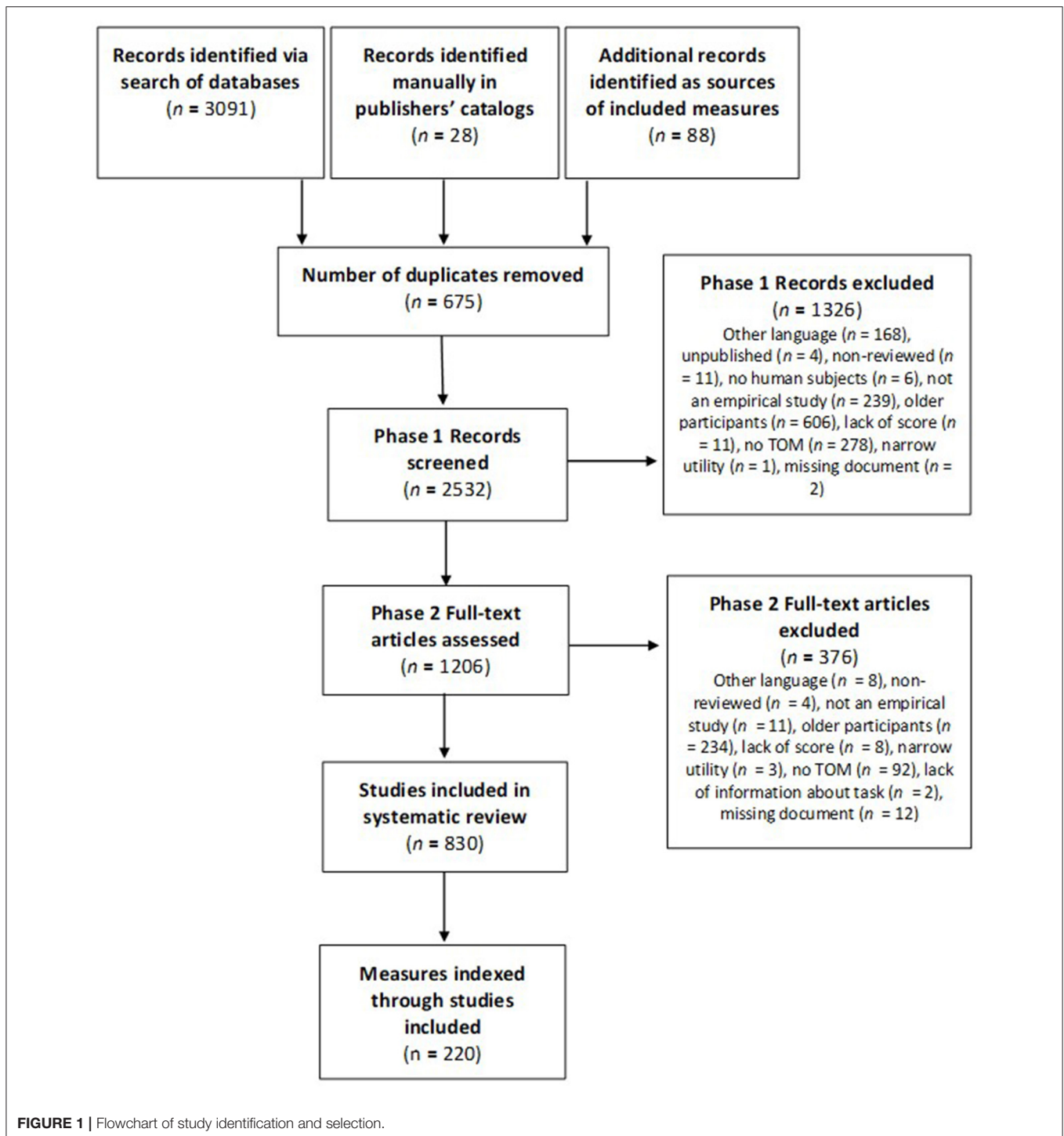
A qualitative content analysis of the measures included was performed by all authors throughout the selection process in order to extract the discrete mental states and social situation understanding that were assessed by the included measures. Seven categories of mental states and social situations were identified across the collection of studies: emotions, desires, intentions, percepts, knowledge, beliefs, and mentalistic understanding of non-literal communication. An eighth category, called "comprehensive measures," was added to represent measures encompassing the understanding of multiple mental states and social situations. These eight TOM categories were therefore used to classify the different measures during data collection.

Data collection was performed by the first three authors using a comprehensive pre-determined form. This form included the following variables related to the measures: category of mental state or social situation assessed, name of measure, author(s), and year of publication, reference(s) of articles that have used the measure, short description, administration format, number of items, scoring options, and administration time. It was also noted which articles provided original psychometric information. The data extraction form also included the following information regarding the participants assessed with the measures: age range of normative population, language(s) spoken, presence of adverse clinical (e.g., hearing impairments or deafness, Williams syndrome), psychological (e.g., anxiety or depression, externalizing behavior problems), or environmental (i.e., low socio-economic status, maltreatment) conditions assessed with the measures.

## RESULTS

### Summary of Main Results and TOM Categories

**Figure 1** illustrates the steps in article selection. A total of 830 studies were included for data extraction. Given the large amount of studies and the numerous variations of the same measures found, a synthesis of the data was performed, which isolated 220 distinct measures and paradigms. Each is presented, along with their characteristics and details of participants that were tested across studies, in tables found in **Appendix II**. **Appendix II** contains eight separate tables according to the main TOM category they refer to: Emotions (Table a; 37 measures), Desires (Table b; 26 measures), Intentions (Table c; 16 measures), Percepts (Table d; 26 measures),



Knowledge (Table e; 25 measures), Beliefs (Table f; 49 measures), Mentalistic understanding of non-literal communication (Table g; 16 measures) and Comprehensive measures (Table h; 25 measures). To further synthesize the results and provide clarity on the content of the tasks, the first seven categories were sub-divided into 39 TOM sub-abilities or sets of abilities assessed in the measures. Category 8, Comprehensive measures,

was subdivided according to the format of the measures (i.e., questionnaires/interviews and direct tests). For example, the Desires category was divided into four sub-abilities: (1) understanding that different people may have discrepant desires, (2) understanding the co-existence of multiple desires at the same time or successively in one person, (3) understanding that people's emotions and actions are influenced by their

desires/preferences, and (4) producing plausible explanations when action contradicts stated desires/preferences.

**Table 2** provides an overview of the results and presents the first seven TOM categories and the 39 TOM sub-abilities, along with an example of a relevant measure and the number of measures and articles that were identified in relation to each sub-ability. **Table 3** presents an overview of the measures included in the Comprehensive measures category. In order to visually represent the organization of the TOM abilities and sub-abilities that emerged from the systematic review, a framework depicting the various types of TOM measures and a related taxonomy was developed and is presented in **Figure 2: Abilities in Theory of Mind Space** (the ATOMS framework).

## Information for Navigating the Results Tables

In the tables (**Appendix II, Tables a–h**), within one TOM sub-ability, measures are presented in alphabetical order according to the first author of the original measure. Articles reporting the use of these measures follow the name of the measure in a numbered format referring to the alphabetical order of authors in the reference list. In addition, within one TOM sub-ability, participants' characteristics are also presented in alphabetical order, when relevant (i.e., languages and adverse conditions). It should be noted that a single article may be cited more than once since it may report the use of more than one TOM measure. Furthermore, measures entailing more than one subtask (i.e., measures from the comprehensive measures category and measures tapping multiple sub-abilities within a specific category) were divided in subtasks and added to the single measures reported, whenever sufficient information was available to do so. Consequently, a single article may be cited as using a comprehensive measure (e.g., Theory of mind scale; Wellman and Liu, 2004) and its subtask (e.g., Content false belief paradigm; Hogrefe et al., 1986; Perner et al., 1987). This procedure for reporting task-related information was applied both to existing tasks embedded in a comprehensive measure (as in the preceding example), as well as, new subtasks created specifically for a comprehensive measure (e.g., Forget stories from the Strange stories; Happé, 1994). In Tables a–h, the column "Availability of psychometric information" informs on the presence (+) or absence (–) of psychometric properties related to a specific measure. When present, the information is then presented in detail in two distinct tables (**Appendix III, Tables i, j**).

When consulting the results tables, readers should be aware of some caveats associated with the data synthesis process. In particular, it is important to note that a specific measure or paradigm may tap more than one TOM category or sub-ability, but for practical reasons, it was placed under the one that was judged to best reflect its measurement scope. For example, the Ella the elephant task (Harris et al., 1989), which captures the emotions associated with false beliefs (e.g., happiness when seeing a can of a preferred beverage, without knowing the content has been replaced by a disliked beverage), was placed in the Beliefs category even though understanding of emotions and desires

are also secondarily involved in the task. Related to this and given the existence of multiple variations of the same paradigms, measures were placed under a common banner when they had strong similarities, even if the authors did not refer directly to the original source. For example, the Ernie test and Linda test, presented by Ford et al. (2012), were referenced under the measure Change-in-location paradigm/Sally and Ann task because they rely on false beliefs associated with the unseen displacement of an object, a paradigm typically attributed to Wimmer and Perner (1983) by most authors. It is also important to note that the original source of a measure may not have been included in the review because of an exclusion criterion (e.g., the original reference for the Emotion Understanding Assessment is in a book; Howlin et al., 1999). In these cases, the source article was not included in the review, but the reference is provided in the tables, beside the name of the measure.

## Measure Characteristics Modes of Presentation

Many different presentation modalities are used across TOM measures, but most rely on direct testing with the child, using read-aloud stories enacted with figurines (19 sub-abilities, e.g., Allen and Kinsey, 2013), or scenarios depicted with pictures (32 sub-abilities, e.g., Galende et al., 2011). Some measures rely on videos (8 sub-abilities, e.g., Mayes et al., 1996), audio-recordings or read-aloud scenarios (21 sub-abilities, e.g., Whitehouse and Hird, 2004), videogames, games or other realistic laboratory situations with the experimenter and/or other persons (14 sub-abilities, e.g., Brown, 2006). Many measures have variations in possible presentation modalities across studies. A good example of this is that all of the references cited in the first part of this section refer to assorted presentation modes of a single measure, the Change-in-location/Sally and Ann task. Most TOM measures use visual support, with few relying solely on verbal information (e.g., Faux pas task used by Hoogenhout and Malcolm-Smith, 2014), and few being entirely non-verbal (e.g., Behavioral re-enactment procedure used by Meltzoff, 1995). Only four measures using a questionnaire format were identified: Everyday mindreading skills and difficulties scale (Peterson et al., 2009), Theory of mind inventory (Hutchins et al., 2008a, 2012), Supplementary social and maladaptive items/*Échelle d'adaptation sociale pour enfants* (Frith et al., 1994) and Children's social understanding scale (Tahiroglu et al., 2014). These are completed by parents and/or a third-party adult, such as a daycare provider or educator.

## Number of Items

The number of items in each measure varies from 1 to 54 in single category measures (Tables a–g) and from 1 to 110 in comprehensive measures (Table h). The number of items administered is highly variable from one study to another. For example, Wellman and Liu's Theory of mind scale (2004) is variably reported as being administered in 3, 4, 5, 6, and 7-item formats, each using a different sampling of items from the original scale (e.g., Davis et al., 2011; Suway et al., 2012; Strasser and del Rio, 2014; Dore and Lillard, 2015). Some authors also

**TABLE 2 |** TOM categories and sub-abilities and associated number of measures and articles.

<b>Mental states and social situations categories</b>	<b>TOM sub-abilities</b>	<b>Examples of measures/paradigms</b>	<b>Number of measures n (%)</b>	<b>Number of articles n (%)</b>
<b>Emotions</b>	1. Typical emotional reactions: Inferring a person's emotional reactions based on situations that typically elicit certain emotions/inferring a preceding event based on a person's emotional reaction	Affective knowledge understanding (Knafo et al., 2009)	19 (8.6%)	66 (8.0%)
	2. Atypical emotional reactions: Inferring or explaining a person's emotional reactions based on situations eliciting emotions that are atypical compared to what is usually expected	Affective perspective-taking (Denham, 1986)	6 (2.7%)	44 (5.3%)
	3. Discrepant emotions: Understanding that people may have discrepant feelings about an event	Affective perspective taking (Borke, 1971; Smith, 1973)	1 (0.5%)	4 (0.5%)
	4. Mixed emotions: Understanding that people may feel mixed emotions or different emotions successively	Mixed emotion understanding task (Gordis et al., 1989)	4 (1.8%)	16 (1.9%)
	5. Hidden emotions: Understanding that other people may hide their emotions	Appearance reality of emotions (Harris et al., 1986)	4 (1.8%)	107 (12.9%)
	6. Moral emotions: Understanding that negative feelings might arise following a reprehensible action	Morality-based emotions (Pons and Harris, 2000)	1 (0.5%)	8 (1.0%)
	7. Emotion regulation: Understanding that others might use strategies to regulate their emotions	Regulation of emotion (Pons and Harris, 2000)	1 (0.5%)	8 (1.0%)
	8. Comprehensive measure involving emotion understanding based on different factors/TOM categories (e.g., desires, beliefs, hiding emotions)	Test of emotion comprehension (Pons and Harris, 2000)	1 (0.5%)	16 (1.9%)
	<b>Emotions category totals</b>		<b>37 (16.8%)</b>	<b>198 (23.9%)</b>
<b>Desires</b>	1. Discrepant desires: Understanding that different people may have discrepant desires	Discrepant desires/Yummy-yucky task (Repacholi and Gopnik, 1997)	10 (4.5%)	130 (15.7%)
	2. Multiple desires: Understanding the co-existence of multiple desires simultaneously or successively in one person	Multiple desires task (Bennett and Galpert, 1993)	5 (2.3%)	5 (0.6%)
	3. Desires influence on emotions and actions: Understanding that people's emotions and actions are influenced by their desires/preferences	Desires task (Wellman and Bartsch, 1988; Wellman and Woolley, 1990)	10 (4.5%)	49 (5.9%)
	4. Desire-action contradiction: Producing plausible explanations when actions contradict stated desires/preferences	Anomalous-desires stories (Colonnesi et al., 2008)	1 (0.5%)	1 (0.1%)
	<b>Desires category totals</b>		<b>26 (11.8%)</b>	<b>178 (21.4%)</b>
<b>Intentions</b>	1. Completion of failed actions: Understanding another person's intent, as demonstrated by completing their failed action	Behavioral re-enactment procedure (Meltzoff, 1995)	1 (0.5%)	12 (1.4%)
	2. Discrepant intentions: Understanding that identical actions/results can be achieved with different intentions	Accidental transgression task (MoToM; Killen et al., 2011)	7 (3.2%)	12 (1.4%)
	3. Prediction of actions: Predicting people's actions based on their intentions	Attention to intention (Phillips et al., 2002)	5 (2.3%)	13 (1.5%)
	4. Intention attribution to visual figures: Tendency to attribute intentions to ambiguous visual figures	Valley task (Castelli, 2006)	1 (0.5%)	1 (0.1%)
	5. Intentions explanations: Producing plausible intention explanations for different types of observed social events	Intentions explanations (Smiley, 2001)	2 (0.9%)	2 (0.2%)
	<b>Intentions category totals</b>		<b>16 (7.3%)</b>	<b>36 (4.3%)</b>
<b>Percepts</b>	1. Simple visual perspective taking: Acknowledging that others have different visual percepts and adopting the visual perspective of another person	Visual perspective taking, Level 1/Picture identification task (Masangkay et al., 1974; Flavell et al., 1981)	15 (6.8%)	80 (9.6%)
	2. Complex visual perspective taking: Adopting another person's visual perspective in tasks demanding complex mental rotation or visualization	Visual perspective taking and spatial construction task (Ebersbach et al., 2011)	9 (4.1%)	14 (1.7%)
	3. Percept-action link: Understanding that other's actions are linked to their visual percepts	Perception based action (Hadwin et al., 1997)	1 (0.5%)	6 (0.7%)

(Continued)



TABLE 2 | Continued

Mental states and social situations categories	TOM sub-abilities	Examples of measures/paradigms	Number of measures n (%)	Number of articles n (%)
Knowledge	4. Auditory perspective taking: Considering the auditory percepts of another person	Auditory perspective taking (Williamson et al., 2015)	1 (0.5%)	1 (0.1%)
	<b>Percepts category totals</b>		<b>26 (11.8%)</b>	<b>97 (11.7%)</b>
	1. Knowledge-pretend play links: Understanding that someone who does not know something exists cannot engage in "pretend play" that incorporates that knowledge	Sarah task (Aronson and Golomb, 1999)	3 (1.4%)	3 (0.4%)
	2. Percepts-knowledge links: Understanding that someone who does not have access to perceptual information (i.e., by looking, hearing, etc.) may not have access to knowledge	See-know task (Pillow, 1989; Ruffman and Olson, 1989)	11 (5.0%)	149 (18.0%)
	3. Information-knowledge links: Understanding that someone who was not informed or is not familiar with something may not know	Awareness of a reader's knowledge task (Peskin et al., 2014)	8 (3.6%)	10 (1.2%)
Beliefs	4. Knowledge-attention links: Understanding that something new is more interesting to someone than something already known	Familiarity-focus of attention (Moll et al., 2006)	2 (0.9%)	3 (0.4%)
	<b>Knowledge category totals</b>		<b>25 (11.4%)</b>	<b>163 (19.6%)</b>
	1. Content false beliefs: Familiar container with an unexpected content: Understanding the false belief held by someone who never opened the container	Content false belief paradigm (Hogrefe et al., 1986; Perner et al., 1987)	4 (1.8%)	414 (49.9%)
	2. Location false beliefs: Unseen change: Understanding the false belief held by someone who did not witness or was not informed of a displacement or change of action	Change-in-location paradigm/Sally-Ann task (Wimmer and Perner, 1983; Baron-Cohen et al., 1985)	7 (3.2%)	396 (47.7%)
	3. Identity false beliefs: Understanding that when something looks/sounds/smells like something else, a person may hold a false belief about its identity	Appearance-reality test (Flavell et al., 1986)	16 (7.3%)	143 (17.2%)
Mentalistic understanding of non-literal communication	4. Second-order belief: Understanding the second-order belief or false belief held by someone who does not know somebody else was informed (e.g., of a misleading identity, a misleading location, etc.)	Ice-cream van test (Perner and Wimmer, 1985)	7 (3.2%)	94 (11.3%)
	5. Beliefs based action/emotions: Predicting another emotions or actions based on their stated beliefs/Inferring another person's belief based on their stated action or emotion	The Tom task (Swettenham, 1996)	8 (3.6%)	154 (18.6%)
	6. Sequence false beliefs: Understanding the false belief created when a predictable sequence of stimuli is broken with the intrusion of an unexpected stimulus	Unexpected outcome (Brambling and Asbrock, 2010)	1 (0.5%)	1 (0.1%)
	7. Comprehensive measures of understanding beliefs	Battery of TOM tasks (Hughes et al., 2000)	6 (2.7%)	20 (2.4%)
	<b>Beliefs category totals</b>		<b>49 (22.3%)</b>	<b>627 (75.5%)</b>
	1. Irony/sarcasm: Understanding that other people may lie in order to be ironic/sarcastic	Lies and jokes task (Sullivan et al., 1995)	6 (2.7%)	19 (2.3%)
	2. Egocentric lies: Understanding that someone may consciously lie in order to avoid a problem or to get its way	Lie stories from the Strange stories (Happé, 1994)	4 (1.8%)	13 (1.6%)
3. White lies: Understanding that someone may lie in order to spare another's feelings	White lie stories from the Strange stories (Happé, 1994)	1 (0.5%)	14 (1.7%)	
4. Involuntary lies: Understanding that someone may tell a "lie" without knowing	Forget stories from the Strange stories (Happé, 1994)	1 (0.5%)	11 (1.3%)	
5. Humor: understanding that someone may tell a "lie" in order to make a joke	Joke stories from the Strange stories (Happé, 1994)	1 (0.5%)	11 (1.3%)	
6. <i>Faux pas</i> : Ability to recognize <i>faux-pas</i> (social gaffe) situations	Recognition of faux pas (Baron-Cohen et al., 1999)	1 (0.5%)	6 (0.7%)	
7. Measures tapping multiple aspects of mentalistic understanding of non-literal communication	Strange stories (Happé, 1994)	2 (0.9%)	22 (2.7%)	
<b>Mentalistic understanding of non-literal communication totals</b>		<b>16 (7.3%)</b>	<b>30 (3.6%)</b>	

Percentages are calculated using the total number of measures (220) and studies (830) included in the review. Some TOM category total of articles are lower than the sum of articles assessing its sub-abilities because each article may assess more than one sub-ability.

indicate that they used only a single task from the Theory of mind scale (e.g., O'Reilly et al., 2014).

**Scoring Options**

Many measures use a simple correct/incorrect scoring scheme (37 sub-abilities) for the child's verbal (e.g., saying where a character will search for an object; Wang et al., 2014) or

behavioral (e.g., giving the experimenter a book he showed a preference for; Laranjo et al., 2010) response to test items. Some measures use a more elaborate scale or coding system (30 sub-abilities) to evaluate children's behavior (e.g., extent to which children adapt their behavior in order for their parent to see an object; Laranjo et al., 2010) or verbal explanation to open-ended questions (e.g., quality of justification when inferring an emotion; Nader-Grosbois et al., 2013). Timing and direction of eye gaze is also used as an indicator of TOM (9 sub-abilities), and assessed using observation coding systems (Poulin-Dubois and Yott, 2014) or eyetracking (Gliga et al., 2014). Of note, from one study to another, there are many adaptations of scoring schemes for the same measure. For example, in two studies using a Change-in-location paradigm/Sally-Ann task to assess false belief understanding, Adrian et al. (2005) asked questions and coded children's verbal answers in a correct/incorrect format, while Senju et al. (2011) coded children's eye movements using an eyetracker.

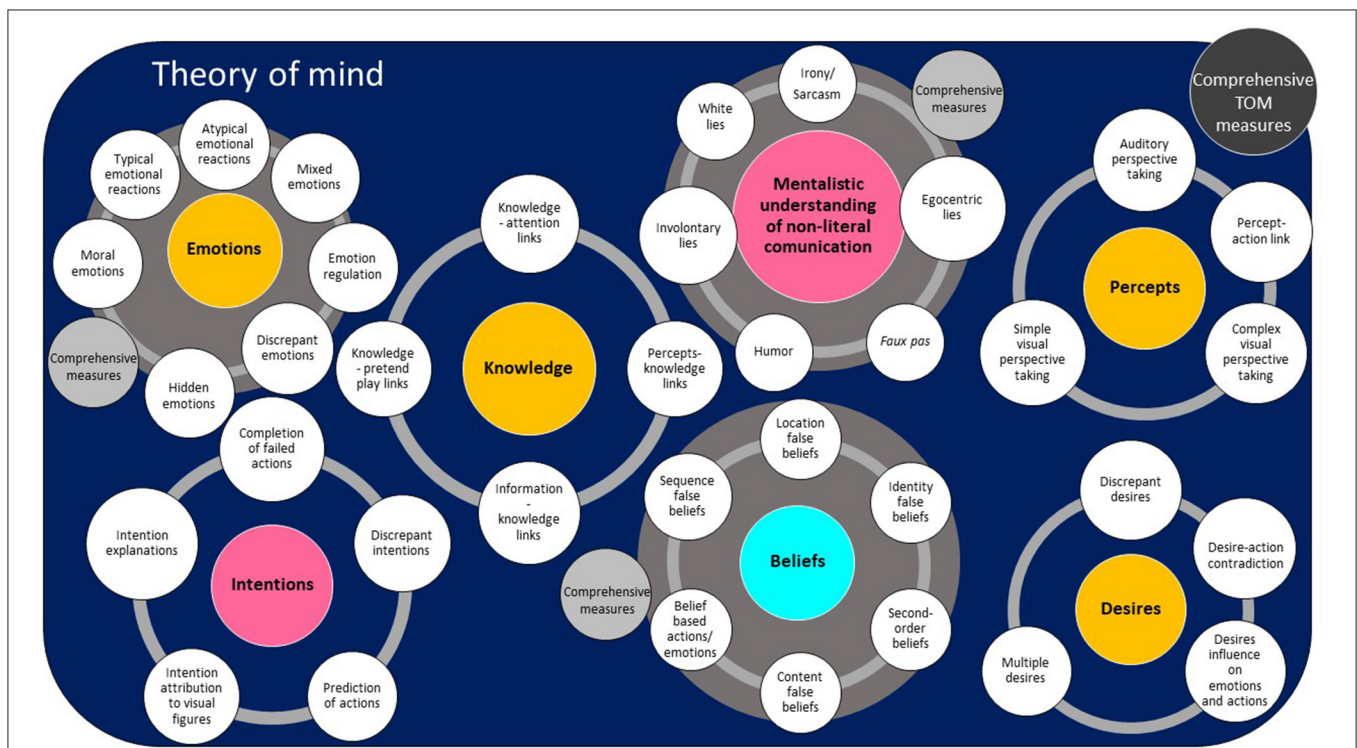
**TABLE 3 |** Comprehensive measures and associated number of measures and articles.

Formats	Examples of measures	Number of measures n (%)	Number of articles n (%)
1. Multiple TOM abilities measured using questionnaires/interviews	Theory of mind inventory (ToMI) (Hutchins et al., 2012)	4 (1.8%)	19 (2.3%)
2. Multiple TOM abilities measured using direct testing	ToM scale (Wellman and Liu, 2004)	21 (9.5%)	183 (22.0%)
<b>Total comprehensive measures</b>		<b>25 (11.4%)</b>	<b>194 (23.4%)</b>

Percentages are calculated using the total number of measures (220) and studies (830) included in the review.

**Administration Time**

While initially extracted from the articles included in the review, administration time was not reported in the final tables



**FIGURE 2 |** ATOMS framework. The ATOMS framework (Abilities in Theory of Mind Space) is a visual representation of the TOM categories and sub-abilities that emerge from the systematic review of TOM measures for young children. Theory of mind space is represented as a large area that includes seven TOM categories of mental states and social situations understanding (colored circles): Intentions, Desires, Emotions, Knowledge, Percepts, Beliefs, and mentalistic understanding of non-literal communication. Thirty-nine specific TOM sub-abilities (white circles) gravitate around the TOM category to which they pertain. When comprehensive measures exist that measure sets of abilities (multiple sub-abilities) for any one TOM categories, these are represented as gray circles. An eighth overall category "Comprehensive TOM measures" includes measures that encompass multiple TOM categories and is represented as a black circle. TOM categories (colored circles) are further represented using three different colors according to the proportion of reviewed studies that measured these types of TOM abilities: the pink circles represent TOM categories measured in <5% of studies, yellow circles represent TOM categories measured in 5–25% of studies, and the blue circle represent the only TOM category (Beliefs) measured in more than 25% of studies.

of results since only a small proportion (5.1%) of authors reported this information. Moreover, it is highly probable that administration time varies substantially from one measure adaptation to another.

### Psychometric Properties

Basic information on internal structure and consistency, inter-rater reliability and test-retest reliability are listed in Tables i and j when available (**Appendix III**), along with the 168 references providing this information (20.2% of included articles). The articles were further qualified as to whether they used an implicit (i.e., non-verbal, indirect and implied cues of children's TOM understanding, such as eye gaze tracking or behavioral observation) or explicit (i.e., direct response provided by the participant, such as verbal responses or pointing to a specific response choice) method for data collection. Forty-one articles (4.9%) provided psychometric properties on implicit methods to measure TOM, using 20 different measures/paradigms. Measures are ordered according to the category of mental state and social situation understanding they pertain to and presented in alphabetical order using the name of the first author of the tool. Articles providing psychometric information are also listed in alphabetical order using first author's name. For many studies, the psychometric data were analyzed using individuals pooled from many age groups and/or adverse conditions. For this reason, the reader is invited to directly consult the studies in order to carefully interpret the data provided. Some studies (e.g., Yagmurlu et al., 2005; Guajardo et al., 2013) report the psychometric properties of aggregates of TOM measures, but these were not included in the tables since they do not refer to one specific measure reviewed. **Table 4** provides an overview of the number of studies providing evidence for or against psychometric validation of four broad categories of indices: internal structure and consistency, inter-rater reliability, test/retest reliability and other psychometric information.

#### *Internal structure and consistency*

Internal consistency refers to the extent to which different items of an assessment tool are inter-correlated, and so refer to the same construct (Terwee et al., 2007). It is recommended to first analyse the structure of the measure, using factor analysis or principal component analysis, to determine/confirm the number of scales before measuring the internal consistency of each scale (Terwee et al., 2007). Of note, hereafter, scaling analyses were not included as formal structure analyses and are instead included in "other psychometric information." Information on internal consistency was found for 37 TOM measures (16.8%) within 72 studies (8.7%). However, only 10 measures also had formal structure analyses (4.5%): three emotions category measures, one Mentalistic understanding of non-literal communication measure and six comprehensive measures. Cronbach alpha is recognized as a good measure of internal consistency and is considered to be adequate when between 0.70 and 0.95 (Terwee et al., 2007). Only four measures had information on their internal structure and their Cronbach's alphas were always between 0.70 and 0.95 across all the studies that provided both structure and consistency information: Children's social understanding scale (Tahiroglu et al., 2014), Theory of

Mind Inventory and Perceptions of Children's Theory of mind inventory and Perceptions of children's theory of mind measure-experimental version (Hutchins et al., 2008b, 2012), TOM task battery (Hutchins et al., 2008b) and "Social meaning scale (SELweb)" (McKown et al., 2016). All the measures were from the comprehensive measures category and all used explicit methods to test TOM.

#### *Reliability*

Inter-rater reliability and test-retest reliability were reported using similar parameters. Weighted Cohen's Kappa coefficient is the most recommended method for reporting the reliability of ordinal measures, whereas an intraclass correlation coefficient is recommended for continuous measures (Terwee et al., 2007). Other inter-rater reliability parameters reported include percentage of agreement and Pearson correlations, which are judged as less adequate measures of reliability (Terwee et al., 2007). *Inter-rater reliability*: Inter-rater reliability was reported for 62 measures (28.2%) within 95 studies (11.4%). Weighted Cohen's Kappa is available for 47 of these measures (21.4%), distributed through all TOM categories. Whenever reported, the Cohen's Kappa coefficients always met the 0.70 minimum standard for reliability, including implicit methods (16 Cohen's Kappa coefficients, reflecting on inter-rater reliability for nine implicit methods/paradigms) (Terwee et al., 2007). *Test-retest reliability*: Test-retest reliability was provided for 18 measures (8.2%) within 15 studies (1.8%), none of which pertained to implicit methods/paradigms. Cohen's Kappa coefficient or intraclass correlation coefficients are available for nine explicit measures (five in the Beliefs category, two in the Comprehensive measures category, one in Percepts category and one in Knowledge category; 4.1%). The 0.70 minimal standard value was attained in all studies reporting this information for three measures: See-know task (Pillow, 1989; Ruffman and Olson, 1989), Message-desire discrepancy (Mitchell et al., 1997) and TOM test (Muris et al., 1999).

#### *Other psychometric information*

Some studies (27 measures, 12.3%; 48 studies, 5.8%) also included other statistics related to a particular measure's psychometric properties. This information is detailed in Tables i, j under "Other psychometric information" and includes, for example, scalability (e.g., Guttman analyses) or construct validity testing, including analyses performed in order to test specific hypotheses regarding the construct validity of the measure (e.g., concurrent and discriminant validity). These additional types of psychometric properties were mostly tested in comprehensive measures (36 out of 48 studies providing specific validity information). In particular, each of the four questionnaires was reported to correlate with TOM scores from direct testing (Hughes et al., 1997; Comte-Gervais et al., 2008; Hutchins et al., 2008a, 2012; Peterson et al., 2009; Houssa et al., 2014; Tahiroglu et al., 2014; Smogorzewska et al., 2019). Among the information retrieved for validity testing, only 10 measures explicitly tested and demonstrated the links between test scores and a measure of social ability: these were all from the comprehensive measures except three tests: Theory of mind inventory (Hutchins et al., 2012), TOM storybooks (Blijd-Hoogewys et al., 2008), TOM test

**TABLE 4 |** Reliability and validity evidence of included TOM measures (number of studies supporting evidence/number of studies less supportive of evidence).

Measures (source author)	Internal consistency	Interrater reliability	Test/retest reliability	Other psychometric information (e.g., scaling analysis, validity testing)
<b>CATEGORY: EMOTIONS UNDERSTANDING</b>				
Affective perspective taking (Cassidy et al., 1992)	2/1	4	0	0
Affective perspective-taking tests (Denham, 1986)	4	1	0	0
Knowledge of emotion cause (Denham et al., 1994)	0	1	0	0
Description of emotional situation (Feshbach and Cohen, 1988)	0	1	0	0
Emotion situation knowledge task (Garner et al., 1994)	1/1	0	0	1
Mixed emotion understanding task (Gordis et al., 1989)	2	1	0	0
Appearance reality of emotions (Harris et al., 1986); Affective false-belief task (Davis, 1998)	0	5	0	0
Emotion understanding assessment (Howlin et al., 1999)	2/1	0	1/0	1
Affective attribution and reasoning task (Iannotti, 1978)	0/1	1	0	0
Test of emotion comprehension (Pons and Harris, 2000)	2/2	0	0	1
Emotion recognition questionnaire (Ribordy et al., 1988)	2/2	0	0	0
<b>CATEGORY: DESIRES UNDERSTANDING</b>				
Diverse desire (Bartsch and Wellman, 1989)	0	1	0	0
Gift task (Flavell, 1968)/Gift selection task (Jin et al., 2017)	0	1	0	0
Discrepant desires Yummy-yucky task (Repacholi and Gopnik, 1997)	0	2	0	0/1
Common and uncommon desires (Rieffe et al., 2001)	1	2	0	0
Desire and intention task (Schult, 2002)	0	2	0	0
Target-hitting game (Schult, 2002)	0	1	0	0
Not own desire tasks (Wellman and Woolley, 1990)	0	4	0	0
Desire task (actions and emotions stories) (Wellman and Woolley, 1990)	1	0	0	0
<b>CATEGORY: INTENTION UNDERSTANDING</b>				
Behavior-, skill-, and awareness-intentionality measures (Astington and Lee, 1991)	0	1	0	0
Visual habituation paradigm (Buresh and Woodward, 2007)	0	2	0	0
Intention and beliefs (Choi and Luo, 2015)	0	1	0	0
Behavioral re-enactment procedure (Meltzoff, 1995)	0	6	0	1/1
Accidental transgression task (MoToM; Killen et al., 2011)	0	1	0	0
Intention task (Phillips and Wellman, 2005)	0	1	0	0/1
Attention to intention (Phillips et al., 2002)	0	1	0	0
<b>CATEGORY: PERCEPTS</b>				
Visual perspective taking and spatial construction task (Ebersbach et al., 2011)	0	1	0	0
Photographers perspective taking (Frick et al., 2014)	1	0	0	0
Penny game task (Gratch, 1964)	2	1	0	1
Perception based action (Hadwin et al., 1997)	0	0	0/1	0
Gaze-following task (Meltzoff and Brooks, 2008)	0	1	0	1
Occluded object task (Moll and Tomasello, 2006)	0	2	0	0
Level-1 perspective taking tasks (Ricard et al., 1999)	0	1	0	0
<b>CATEGORY: KNOWLEDGE</b>				
Cognitive perspective taking (Brice and Torney-Purta, 1981)	0/1	0	0	0
Cognitive perspective taking (Flavell, 1968)	0	1	0	0
Familiarity-focus of attention (Moll et al., 2006)	0	1	0	0
Knowledge theory of mind task (Moll and Tomasello, 2007)	0	2	0	0
See-know task (Pillow, 1989; Ruffman and Olson, 1989)	0	5	1	0
Hide an object (Viranyi et al., 2006)	0	1	0	0
<b>CATEGORY: BELIEFS UNDERSTANDING</b>				
Deceptive contents false-belief task (Bartsch and Wellman, 1989)	1/1	2	0	1
Picture false-belief task (Callaghan et al., 2012)	0	1	0	0
Lexical ambiguity (Carpendale and Chandler, 1996)	0	1	0	0

(Continued)

TABLE 4 | Continued

Measures (source author)	Internal consistency	Interrater reliability	Test/retest reliability	Other psychometric information (e.g., scaling analysis, validity testing)
Doodle task (Chandler and Helm, 1984; Hughes et al., 1997)	0	2	0	0/1
Appearance-reality tasks (Flavell et al., 1986)	0	1	0	0
Ella the elephant or Emotion false belief task (Harris et al., 1989)	0	0	0/1	0
Content false belief paradigm (Hogrefe et al., 1986; Perner et al., 1987)	2	13	0/1	1
Battery of TOM tasks (Hughes et al., 2000)	0/2	1	0/1	0
ToM task (Kim and Phillips, 2014)	1	0	0	0
Message-desire discrepancy (Mitchell et al., 1997)	0	0	1	0
False-belief suspense (Moll et al., 2016)	0	1	0	0
Ice-cream van test (Perner and Wimmer, 1985)	0	1	0/1	0
False belief story (Riggio and Cassidy, 2009)	0	1	0	0
Birthday puppy (Sullivan et al., 1995)	0	1	0	0
Granddad story, Window story or Tom's crayon (Sullivan et al., 1995; Astington et al., 2002)	1	2	0/2	0
Ambiguity task (Taylor et al., 1988)	0	2	0	0
False-belief explanation task (de Villiers and de Villiers, 2000)	0/1	0	0	0
Belief tasks (Wellman and Bartsch, 1988)	0	7	0	0
Change-in-location paradigm (Wimmer and Perner, 1983)/Sally-Ann task (Baron-Cohen et al., 1985)	1/2	22	1/2	2/5
<b>CATEGORY: MENTALISTIC UNDERSTANDING OF NON-LITERAL COMMUNICATION</b>				
Irony task (Filippova and Astington, 2008)	0	1	0	0
Joke stories from the Strange stories (Happé, 1994)	0	1	0	0
Sarcasm stories from the Strange stories (Happé, 1994)	0	2	0	0
Strange stories (Happé, 1994)	2/1	6	0	0
White lies stories from the Strange stories (Happé, 1994)	0	1	0	0
Recognition of faux pas (Baron-Cohen et al., 1999)	1/1	1	0	1
<b>CATEGORY: COMPREHENSIVE MEASURES (DIRECT TESTS)</b>				
ToM storybooks (Blijd-Hoogewys et al., 2008)	1/1	1	1	2
Psychological explanation task (Colonnesi et al., 2008)	0	1	0	0
Comic strip task (Cornish et al., 2010)	0/1	0	0	1
Perspective taking task (Edelstein et al., 1984)	0	1	0	1
TOM task battery (Hutchins et al., 2008b)	2	2	2	1
Theory of mind subtest from a developmental neuropsychological assessment (NEPSY-II; Korkman et al., 2007)	1	1	1	1
Perspective-taking (Krcmar and Vieira, 2005)	1	1	0	0
Pragma test (Loukusa et al., 2014)	0	1	0	1
Social meaning scale from the SELweb (McKown et al., 2016)	1	0	0/1	1
TOM test (Muris et al., 1999)	3/1	2	1	1
Perspective-taking tasks (Oppenheimer and Rempt, 1986)	1	0	0	0
Theory of mind test (Pons and Harris, 2000)	0/1	0	0	0
Explanation of action task (Tager-Flusberg and Sullivan, 1994)	0	1	0	0
ToM scale (Wellman and Liu, 2004)	9/5	9	0	15
<b>CATEGORY: COMPREHENSIVE MEASURES (QUESTIONNAIRES)</b>				
Supplementary social and maladaptive items/ <i>Échelle d'adaptation sociale pour enfants</i> (Frith et al., 1994)	1	1	0	2
Theory of mind inventory and Perceptions of children's theory of mind measure-experimental version (Hutchins et al., 2008a, 2012)	4	0	3	5
Everyday mindreading skills and difficulties scale (Peterson et al., 2009)	1	0	0	1
Children's social understanding scale (Tahiroglu et al., 2014)	3	0	1/1	2

Evidence was generally determined as a value over 0.70 for internal consistency, inter-rater reliability, test/retest and intra-rater (other psychometric information) coefficients, or as % agreement over 80%, and as performed scaling analyses or explicitly tested convergence validity or replicability leading to positive results (other psychometric information). Included measures that are not part of this table had no studies reporting on their psychometric properties. Please see tables i and j (Appendix III) for more detailed psychometric information.

(Muris et al., 1999), TOM task battery (Hutchins et al., 2008b), Theory of mind scale (Wellman and Liu, 2004), Social meaning scale from the SEL web (McKown et al., 2016), Children's social understanding scale (Tahiroglu et al., 2014), Emotion situation knowledge task (Garner et al., 1994), Emotion understanding assessment (Howlin et al., 1999) and Recognition of faux pas (Baron-Cohen et al., 1999). Other important information presented in this section pertains to results from replicability testing: six studies reported independent results replication attempts using five TOM measures, including different variations in their modes of presentation and scoring methods. Most of those studies targeted implicit measures and were not or only partially able to replicate the past results. It is important to note that only articles providing clear objectives to test the validity or reliability of a measure were listed in the tables. However, multiple other articles may provide implicit cues regarding the validity of a measure, such as correlations with other relevant constructs.

## Participant Characteristics

### Languages

While the majority of study samples were comprised exclusively of English-speaking participants (597 studies, 71.9%), some measures were also administered to children speaking 39 other languages (233 studies, 28.1%).

### Age of Typically Developing Children Assessed

While this review specifically aimed to retrieve measures used with young children, typically developing children and adolescents across the pediatric range have also been tested using the measures identified. The youngest typically developing participants reported were 6 months old (Sodian et al., 2016) and some studies included both children and adults (e.g., Reed, 1994; Hirai et al., 2013). Infants have been tested using Intentions (age range: 6 months–17 years old), Percepts (age range: 11 months–40 years old), Desires (age range: 12 months–29 years old), Beliefs (age range: 12 months–92 years old) and Knowledge (age range: 17 months–16 years old) categories of TOM, whereas other categories are limited to older participants (Emotions: 23 months–15 years old; Mentalistic understanding of non-literal communication: 36 months–16 years old).

### Adverse Conditions

In addition to using the measures with typically developing participants, many studies report on their use in children, adolescents or adults with medical (e.g., deafness), psychological (e.g., anxiety or mood disorders), or environmental (i.e., low SES and maltreatment) adverse conditions (236 studies, 28.4%). Thirty different conditions were documented throughout the measures reviewed (Figure 3). The most frequently studied conditions were autism spectrum disorders (118 studies, 14.2%), low socio-economic status (37 studies, 4.5%), hearing impairments and deafness (28 studies, 3.4%), intellectual disability and developmental delay (26 studies, 3.1%), and language impairments (20 studies, 2.4%).

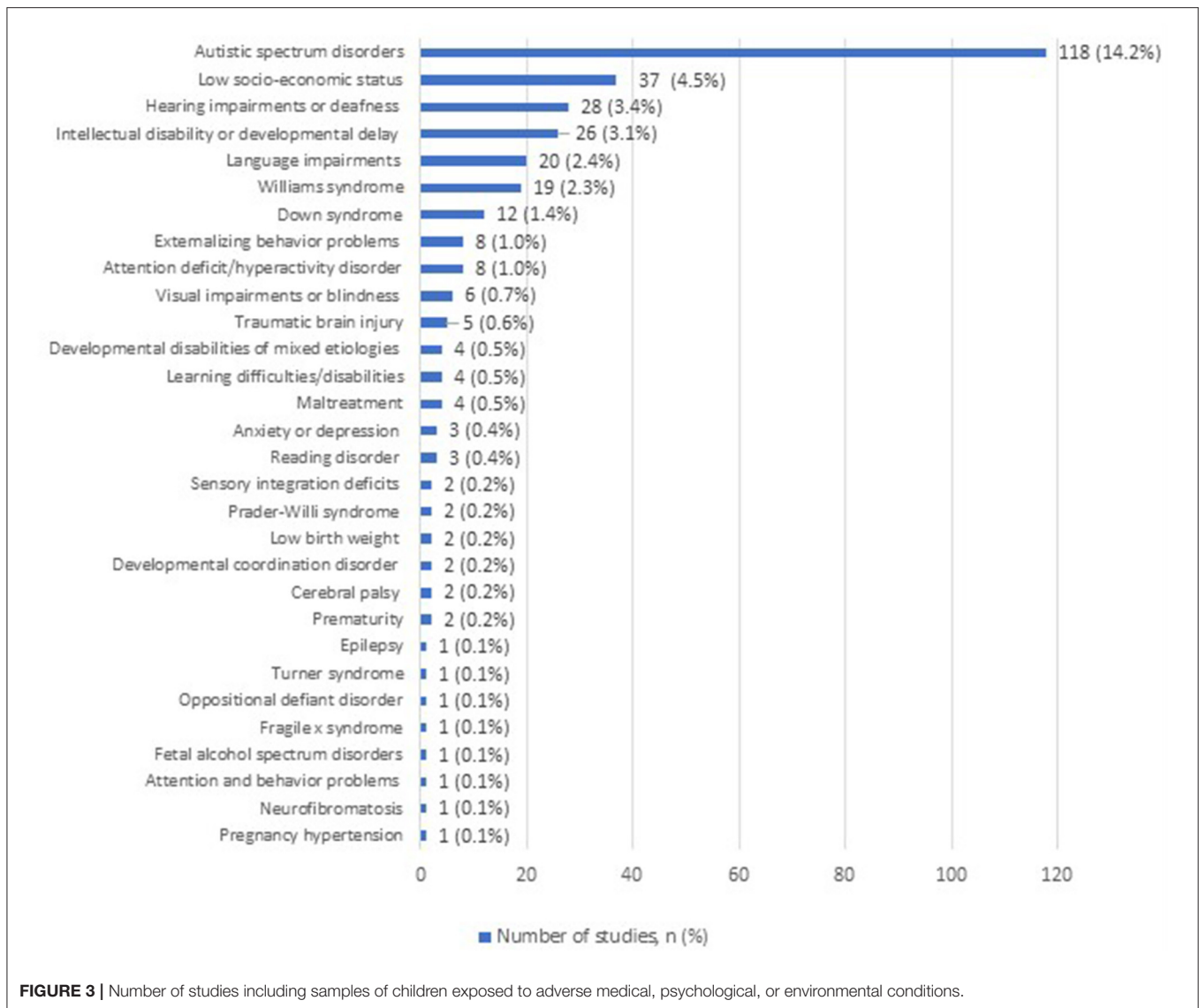
## DISCUSSION

Peer-reviewed literature and relevant test publishers' catalogs were systematically screened in order to generate an inventory of existing TOM measures that have been used with children under 6 years of age. A total of 220 measures, identified through 830 studies, were found to assess the understanding of seven different categories of mental states and social situations: Emotions, Desires, Intentions, Percepts, Knowledge, Beliefs, and Mentalistic understanding of non-literal communication. These were further divided into 39 distinct TOM sub-abilities that have been studied in infants, toddlers and preschoolers. In addition, an eighth category, Comprehensive measures, is comprised of tools assessing multiple categories. To our knowledge, this is the first comprehensive systematic review conducted to document of TOM measures for individuals of any age. This research extends the findings of previous non-systematic literature reviews in other populations (Sprung, 2010; Henry et al., 2015; Hayward and Homer, 2017) and of a systematic review targeting specifically comprehensive and validated TOM measures in preschool children (Ziatabar Ahmadi et al., 2015), and provides a more complete picture of existing TOM assessment methods that can be used with children under the age of six. Information gleaned from the measures and from the review provides an opportunity to identify some of the challenges and future directions associated with TOM assessment.

## Contributions, Challenges, and Possibilities in Relation to TOM Assessment

### Diversity of TOM Abilities

In the last 36 years, studies have focused primarily on TOM abilities related to understanding of Beliefs (75.5% of studies), with fewer studies focussing on other aspects of TOM, such as the understanding of Emotions (23.9%), Desires (21.4%), Intentions (4.3% of studies), and Knowledge (19.6% of studies). However, it appears that an increasing number of studies use Comprehensive measures (23.4%) that tap more than one category of mental states and social situation understanding. These findings align with efforts to diversify sampling of TOM skills when assessing social cognition, in order to better capture its complex nature (Carlson et al., 2013; Hiller et al., 2014; Ziatabar Ahmadi et al., 2015). To this effect, Hiller et al. (2014) underscore the idea that isolated tests do not capture the rich manifestations of TOM abilities, limit the contributions of informative longitudinal assessment, and are an obstacle to understanding TOM development (Hiller et al., 2014). Social cues are among the most complex stimuli that the human brain has to process and are subject to both experiential and environmental influences; measures of social cognition should therefore reflect the complex nature of social stimuli and situations (Beauchamp, 2017). The measurement of more diverse TOM abilities, rather than a narrow focus on false belief understanding, could help enhance external validity, which was rarely tested in the studies



included in this review, and has not typically been supported in other research (Happé et al., 2017).

### Applications and Contributions of the ATOMS Framework

This review led to the elaboration of a new TOM taxonomy, the ATOMS framework (7 categories, 39 sub-abilities). While the primary goal of this classification was to facilitate synthesis and to structure the presentation of a substantial amount of data, the framework also provides an opportunity to reflect on theoretical, methodological and clinical challenges pertaining to TOM. At a theoretical level, the ATOMS classification highlights the need to better conceptualize TOM as a construct. To date, theoretical models mostly aim to explain the links between TOM and other socio-cognitive constructs, such as empathy, emotion recognition and pretend play (Leslie, 1987; Tager-Flusberg and Sullivan, 2000; Abu-Akel and Shamay-Tsoory, 2011; Bird and Viding, 2014; Happé and Frith, 2014; Westby, 2014;

Asakura and Inui, 2016; Happé et al., 2017), but give few details on the make-up of TOM itself. The lack of theoretical structure and shared taxonomy in TOM definitions and its underlying composition impedes our ability to fully integrate TOM in a coherent and comprehensive framework linking it to various socio-cognitive abilities, a pervasive issue observed across the domain of social cognition (Beauchamp, 2017; Happé et al., 2017). The ATOMS framework provides structure for detailing TOM sub-components and for associating them with a nomenclature that could be applied to other work.

This classification may also contribute to guiding the development and interpretation of more comprehensive research protocols and clinical evaluations. The inventory may help enrich TOM evaluation by increasing and diversifying the TOM abilities that are targeted. It could also promote the creation of more comprehensive assessment tools, inspired by the multiple skills composing TOM and the variety of existing measurement methods highlighted in this review.

In research and clinical settings, measures could be more precisely chosen and interpreted to target specific TOM abilities (Happé et al., 2017).

### Diversity of Measurement Methods

This review highlights the creativity drawn on by those who develop new TOM measures, as reflected in the large variety of modes of presentation and administration: scenarios enacted directly with children and/or their entourage, scenarios enacted with the support of figurines, pictures, videos or audio-recordings, games played between the experimenter and the child, videogames, and so on. Measures have also been created or adapted for use with different populations: 40 different languages and 30 distinct adverse conditions are reported (e.g., hearing impairments, visual impairments, autism spectrum disorders).

Given that many other social measures have been limited to questionnaires (Crowe et al., 2011), it is somewhat surprising that only four adult-report questionnaires were found that measure TOM in young children, and these were only used in 2.4% of studies. Direct testing with children is therefore prominent in TOM research and represents a strength of the field, given that direct, laboratory testing provides an explicit opportunity for observing children's responses and may reduce bias associated with parental reports. However, sole reliance on direct testing may also have limits, because it depends on a single context (laboratory) and a single source of information (child) (Carlson et al., 2013). Given that triangulation of data is of importance in clinical (American Psychiatric Association, 2013; American Educational Research Association, A. P. A., and National Council on Measurement in Education, 2014) and research settings (Tashakkori and Teddlie, 2010), and that TOM abilities exhibited in the laboratory are not consistently applied in everyday life (Happé et al., 2017), collecting third party observations on children's natural functioning in social environments via questionnaires or interviews could provide additional information on the behavioral manifestations of TOM. Moreover, initial psychometric data on these questionnaires supports their convergent construct validity. Specifically, each of the four questionnaires was reported to correlate with TOM direct testing scores (Hughes et al., 1997; Comte-Gervais et al., 2008; Hutchins et al., 2008a, 2012; Peterson et al., 2009; Houssa et al., 2014; Tahiroglu et al., 2014). Other promising avenues to conduct ecological evaluation are related to the use of virtual reality and naturalistic, real-world observations of children's behavior, approaches that have seldom been used to date, but that may become more feasible as technology advances and with greater awareness of the importance of the use of real social stimuli in social cognitive assessment (Beauchamp, 2017).

### Enrichment of Measurement Tools

This literature review portrays the structure of TOM measures used to date. Many measures reviewed here rely on only one or two test items when measuring a specific ability, essentially creating a "pass or fail" situation for the examinee,

a problem that has also been raised by others (Cutting and Dunn, 1999; Garner et al., 2005). Such tools offer little score variation and sensitivity to qualify participants' social competence. As with other cognitive functions, TOM should be situated on a continuum and not treated dichotomously (capable or incapable). The need to collect a sample of items large enough to represent any psychological construct is a well-recognized issue in the establishment of adequate content validity and reliability (Slick, 2006; American Educational Research Association, A. P. A., and National Council on Measurement in Education, 2014). The numerous measures listed in this review provide several examples of tests and test items that could be used in order to enrich the evaluation on any TOM category or sub-ability.

### Standardization of TOM Assessment

There is a sizeable number of variations in single tasks across studies. Synthesizing the data extracted in this review presented a significant challenge, owing to the numerous "free" adaptations of unique measures found in the literature. This added a layer of complexity when deciding whether an adaptation of a measure or paradigm should be seen as distinct from the original or not. The wide assortment of TOM measures leads to poor comparability across studies (Hiller et al., 2014) and can be detrimental to the reliability of results (Slick, 2006). For example, success on false belief paradigms may vary as a function of seemingly superficial aspects of the task, such as the type of material used (e.g., is it familiar to the child or new?; Adrien et al., 1995; Cassidy, 1998), the characters presented (e.g., are they real people or figurines?; Battacchi et al., 1997), and subtle differences in language used to question the child (e.g., positive or negative sentence?; Abu-Akel and Bailey, 2001; Geangu, 2002). These task variations constitute a challenge for researchers and clinicians seeking to identify the best measures among all existing task variations found in the literature.

### Psychometric Properties of TOM Measures

This systematic review confirms some of the critiques that have been raised regarding TOM psychometry (Hutchins et al., 2008a; Hiller et al., 2014; Ziatabar Ahmadi et al., 2015). Notably, insufficient TOM measures have empirically validated psychometric properties: internal structure or internal consistency information was available for 37 measures, inter-rater reliability information was available for 62 measures, test-retest reliability was available for 18 measures, other psychometric information, including validity hypothesis testing, was available for only 27 measures. While presenting interesting inter-rater reliability data, implicit methods to measure TOM failed to provide any information on test-retest reliability and are challenged by independent replication studies suggesting globally poor replicability. It should be noted that the current study was not intended to comprehensively review and critique psychometric properties of TOM measures to provide guidelines for measure selection. This objective would require a specific methodology, including assessing study quality and reporting



separate psychometric properties for different versions of the same tasks. The readers are thus invited to exercise caution when interpreting the psychometric data included in this review. Nevertheless, the summary tables included here provide basic information to begin a more detailed search of published psychometric properties for TOM measures. While pursuing such a search, readers should exert their judgment regarding the methodological quality of the validation studies, since the same psychometric property may be more or less powerful depending on study design (e.g., number of participants) and measure characteristics (e.g., number of items). Guidelines for evaluating the quality of tools, such as those published by Terwee et al. (2007), may be useful as they list psychometric properties and gold standard validation methodologies. The psychometric properties reported are likely only to reflect the properties of the specific version of the measure used in a particular study, and not necessarily other adaptations of the measure. Finally, lack of psychometric properties for a specific measure in the results tables does not necessarily reflect disregard of their importance on the part of the authors; some describe psychometric properties of aggregates of single measures (e.g., Yagmurlu et al., 2005; Guajardo et al., 2013), and these were not included in the current review since they did not refer to a specific measure.

## Limitations

The results of this systematic review should be interpreted in the context of certain limitations. First, given the large amount of search results obtained via electronic databases, publishers' catalogs and other sources (3,207 records), additional searches of the gray literature, such as screening of the references in the 830 articles was not performed, even though it is possible that this may have revealed additional measures or additional information on the measures listed herein (Moher et al., 2015). Second, despite the numerous search terms used, the selection of keywords and truncations to capture related terms, and the large amount of measures and articles found, the search strategy failed to retrieve a few pertinent articles that fit the inclusion criteria (e.g., Chen and Lin, 1994; Meltzoff, 1995; Tardif et al., 2004). This is likely due to a lack of common vocabulary in the field, with authors using different terms to refer to similar constructs somewhat interchangeably (i.e., "mentalizing," "mind-reading," and "theory of mind"; Happé et al., 2017). Third, the theoretical model selected to define TOM (SOME model; Bird and Viding, 2014) necessarily determined the inclusion and exclusion criteria for the review. As such, the review may have excluded measures that would have been identified as TOM tools using other models/definitions. In particular, implicit measures of the ability to infer mental states in others, often used with children under 2 years, were only partially captured (see Scott and Baillargeon, 2017 for a review of non-traditional and implicit methods used to measure TOM). Moreover, measures that were judged to primarily assess classification of affective cues (e.g., Reading the mind in the eyes task; Baron-Cohen et al., 1997) and cooperation and competition tasks were not included (e.g., Window task;

Russell et al., 1991), nor were those that document the use (e.g., number of mental state terms used by the child; Internal state language questionnaire, Bellagamba et al., 2014) and understanding (e.g., understanding the difference between the words "know" and "believe"; Certainty task, Adrian et al., 2007) of mental state language, or children's verbal explanations when faced with TOM paradigms (e.g., Peskin and Astington, 2004; Veneziano and Christian, 2006). Fourth, this review did not cover "control tasks," that is, tasks that match TOM tasks in terms of cognitive demands and modes of presentation, but that do not require mental state inferences. For example, there exists a control task for the change-in-location paradigm called the Natural false sign location (e.g., Lackner et al., 2012). The use of control tasks is increasingly recommended in order to take into account the confounding effect of general cognitive abilities and to identify specific social cognition impairments (Henry et al., 2016).

## Conclusions

This systematic review of TOM measures destined for young children identified 830 articles and 220 measures published in the last 36 years that have been administered in 40 different languages and in the context of 30 different medical, psychological and environmental adverse conditions, confirming the preponderance of TOM in many domains of research and practice. The detailed inventory of TOM measures is accompanied by a TOM taxonomy (ATOMS), which presents categories of mental states and social situation understanding that have been used in published research with young children. The findings associated with the review underscore a number of important challenges in TOM assessment. Given that interest in TOM and associated social cognitive constructs is pervasive across developmental psychology, neuropsychology, social psychology, educational psychology and social neuroscience research, and that the need to assess and intervene within these domains is now recognized clinically (Steerneman et al., 1996; Sprung, 2010; Hoddenbach et al., 2012; Lecce et al., 2014; Henry et al., 2016; Beauchamp, 2017), this inventory of TOM measures contributes to both fundamental science and clinical practice.

## DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

CB, CG, and MB contributed to the conception and design of the study. CB, ÉL, and CG collected and analyzed the data. CB wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version. MB supervised the study.

## FUNDING

This work was supported by the *Fonds de Recherche du Québec-Société et Culture* (grant number 198516) and a *Fonds de Recherche du Québec-Santé* fellowship (grant number 32680).

## ACKNOWLEDGMENTS

The authors would like to acknowledge the help of Dominic Desaulniers, psychology librarian at the University of Montreal,

in the creation of the search strategy, and Geneviève Morin, Lara-Kim Huynh and Pascale Mackay, for their assistance in the preparation of tables.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02905/full#supplementary-material>

## REFERENCES

- Abu-Akel, A., and Bailey, A. L. (2001). Indexical and symbolic referencing: what role do they play in children's success on theory of mind tasks? *Cognition* 80, 263–281. doi: 10.1016/S0010-0277(00)00149-9
- Abu-Akel, A., and Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia* 49, 2971–2984. doi: 10.1016/j.neuropsychologia.2011.07.012
- Adrian, J. E., Clemente, R. A., and Villanueva, L. (2007). Mothers' use of cognitive state verbs in picture-book reading and the development of children's understanding of mind: a longitudinal study. *Child Dev.* 78, 1052–1067. doi: 10.1111/j.1467-8624.2007.01052.x
- Adrian, J. E., Clemente, R. A., Villanueva, L., and Rieffe, C. (2005). Parent-child picture-book reading, mothers' mental state language and children's theory of mind. *J. Child Lang.* 32, 673–686. doi: 10.1017/S0305000905006963
- Adrien, J., Rossignol, C., Barthelemy, C., and Jose, C. (1995). Development and functioning of a "theory of mind" in autistic and normal children. *Approch. Neuropsychol. Apprent. l'Enfant* 7, 188–196.
- Allen, J. R., and Kinsey, K. (2013). Teaching theory of mind. *Early Educ. Dev.* 24, 865–876. doi: 10.1080/10409289.2013.745182
- American Educational Research Association, A. P. A., and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: AERA Publications.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington, DC: American Psychiatric Publishing.
- Anastassiou-Hadjicharalambous, X., and Warden, D. (2008). Cognitive and affective perspective-taking in conduct-disordered children high and low on callous-unemotional traits. *Child Adolesc. Psychiatry Ment. Health* 2:16. doi: 10.1186/1753-2000-2-16
- Aronson, J. N., and Golomb, C. (1999). Preschoolers' understanding of pretense and presumption of congruity between action and representation. *Dev. Psychol.* 35, 1414–1425. doi: 10.1037/0012-1649.35.6.1414
- Asakura, N., and Inui, T. (2016). A bayesian framework for false belief reasoning in children: a rational integration of theory-theory and simulation theory. *Front. Psychol.* 7:2019. doi: 10.3389/fpsyg.2016.02019
- Astington, J. W., and Lee, E. (1991). "What do children know about intentional causation?," in *Paper Presented at the Biennial Meeting of the Society for Research in Child Development* (Seattle, WA).
- Astington, J. W., Pelletier, J., and Homer, B. (2002). Theory of mind and epistemological development: the relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas Psychol.* 20, 131–144. doi: 10.1016/S0732-118X%2802%2900005-3
- Baron-Cohen, S., Jolliffe, T., Mortimore, C., and Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *J. Child Psychol. Psychiatry Allied Discipl.* 38, 813–822. doi: 10.1111/j.1469-7610.1997.tb01599.x
- Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8
- Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., and Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *J. Autism Dev. Disord.* 29, 407–418. doi: 10.1023/A:1023035012436
- Bartsch, K., and Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Dev.* 60, 946–964. doi: 10.2307/1131035
- Battacchi, M. W., Celani, G., and Bertocchi, A. (1997). The influence of personal involvement on the performance in a false belief task: a structural analysis. *Int. J. Behav. Dev.* 21, 313–329. doi: 10.1080/016502597384893
- Baurain, C., and Nader-Grosbois, N. (2013). Theory of mind, socio-emotional problem-solving, socio-emotional regulation in children with intellectual disability and in typically developing children. *J. Autism Dev. Disord.* 43, 1080–1097. doi: 10.1007/s10803-012-1651-4
- Beauchamp, H. M. (2017). Neuropsychology's social landscape: common ground with social neuroscience. *Neuropsychology* 31, 981–1002. doi: 10.1037/neu0000395
- Beauchamp, H. M., and Anderson, V. (2010). SOCIAL: an integrative framework for the development of social skills. *Psychol. Bull.* 136, 39–64. doi: 10.1037/a0017768
- Bellagamba, F., Laghi, F., Lonigro, A., Pace, C. S., and Longobardi, E. (2014). Concurrent relations between inhibitory control, vocabulary and internal state language in 18- and 24-month-old Italian-speaking infants. *Eur. J. Dev. Psychol.* 11, 420–432. doi: 10.1080/17405629.2013.848164
- Bellerose, J., Beauchamp, M. H., and Lassonde, M. (2011). "New insights into neurocognition provided by brain mapping: social cognition and theory of mind," in *New Insights Into Neurocognition Provided by Brain Mapping: Social Cognition and Theory of Mind*, ed H. Duffau (Paris: Springer Verlag), 181–192. doi: 10.1007/978-3-7091-0723-2\_14
- Bellerose, J., Bernier, A., Beaudoin, C., Gravel, J., and Beauchamp, M. H. (2017). Long-term brain-injury-specific effects following preschool mild TBI: a study of theory of mind. *Neuropsychology* 31, 229–241. doi: 10.1037/neu0000341
- Benarous, X., Guilé, J.-M., Consoli, A., and Cohen, D. (2015). A systematic review of the evidence for impaired cognitive theory of mind in maltreated children. *Front. Psychiatry* 6:108. doi: 10.3389/fpsyg.2015.00108
- Bennett, M., and Galpert, L. (1993). Children's understanding of multiple desires. *Int. J. Behav. Dev.* 16, 15–33. doi: 10.1177/016502549301600102
- Biedermann, F., Frajo-Apor, B., and Hofer, A. (2012). Theory of mind its relevance in schizophrenia. *Curr. Opin. Psychiatry* 25, 71–75. doi: 10.1097/YCO.0b013e3283503624
- Binnie, L. M. (2005). TOM goes to school: theory of mind understanding and its link to schooling. *Educ. Child Psychol.* 22, 81–93.
- Bird, G., and Viding, E. (2014). The self to other model of empathy: providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neurosci. Biobehav. Rev.* 47, 520–532. doi: 10.1016/j.neubiorev.2014.09.021
- Blakemore, S. J. (2008). The social brain in adolescence. *Nat. Rev. Neurosci.* 9, 267–277. doi: 10.1038/nrn2353
- Blakemore, S. J. (2012). Imaging brain development: the adolescent brain. *Neuroimage* 61, 397–406. doi: 10.1016/j.neuroimage.2011.11.080
- Blijd-Hoogewys, E. M., van Geert, P. L., Serra, M., and Minderaa, R. B. (2008). Measuring theory of mind in children. Psychometric properties of the TOM storybooks. *J. Autism Dev. Disord.* 38, 1907–1930. doi: 10.1007/s10803-008-0585-3
- Bora, E., and Köse, S. (2016). Meta-analysis of theory of mind in anorexia nervosa and bulimia nervosa: a specific impairment of cognitive perspective taking in anorexia nervosa? *International J. Eating Disord.* 49, 739–740. doi: 10.1002/eat.22572

- Bora, E., and Meletti, S. (2016). Social cognition in temporal lobe epilepsy: A systematic review and meta-analysis. *Epilepsy Behav.* 60, 50–57. doi: 10.1016/j.yebeh.2016.04.024
- Bora, E., and Pantelis, C. (2016). Meta-analysis of social cognition in attention-deficit/hyperactivity disorder (ADHD): comparison with healthy controls and autistic spectrum disorder. *Psychol. Med.* 46, 699–716. doi: 10.1017/S0033291715002573
- Bora, E., Yücel, M., and Pantelis, C. (2009). Theory of mind impairment: a distinct trait-marker for schizophrenia spectrum disorders and bipolar disorder? *Acta Psychiatr. Scand.* 120, 253–264. doi: 10.1111/j.1600-0447.2009.01414.x
- Borke, H. (1971). Interpersonal perception of young children: egocentrism or empathy? *Dev. Psychol.* 5, 263–269. doi: 10.1037/h0031267
- Brambling, M., and Asbrock, D. (2010). Validity of false belief tasks in blind children. *J. Autism Dev. Disord.* 40, 1471–1484. doi: 10.1007/s10803-010-1002-2
- Brice, P., and Torney-Purta, J. (1981). *Play and Perspective-taking in Preschool Children: a Study Using Cartoon and Puppet Measures*. University of Illinois, Chicago Circle.
- Brown, C. S. (2006). Bias at school: Perceptions of racial/ethnic discrimination among Latino and European American children. *Cogn. Dev.* 21, 401–419. doi: 10.1016/j.cogdev.2006.06.006
- Brune, M. (2001). Social cognition and psychopathology in an evolutionary perspective. Current status and proposals for research. *Psychopathology* 34, 85–94. doi: 10.1159/000049286
- Brune, M. (2005). Emotion recognition, ‘theory of mind,’ and social behavior in schizophrenia. *Psychiatry Res.* 133, 135–147. doi: 10.1016/j.psychres.2004.10.007
- Buresh, J. S., and Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition* 104, 287–314. doi: 10.1016/j.cognition.2006.07.001
- Byom, L. J., and Turkstra, L. (2012). Effects of social cognitive demand on Theory of Mind in conversations of adults with traumatic brain injury. *Int. J. Lang. Commun. Disord.* 47, 310–321. doi: 10.1111/j.1460-6984.2011.00102.x
- Cacioppo, J. T. (2002). Social neuroscience: understanding the pieces fosters understanding the whole and vice versa. *Am. Psychol.* 57, 819–831. doi: 10.1037/0003-066X.57.11.819
- Callaghan, T. C., Rochat, P., and Corbit, J. (2012). Young children’s knowledge of the representational function of pictorial symbols: development across the preschool years in three cultures. *J. Cogn. Dev.* 13, 320–353. doi: 10.1080/15248372.2011.587853
- Carlson, S. M., Koenig, M. A., and Harms, M. B. (2013). Theory of mind. *WIREs Cogn. Sci.* 4, 391–402. doi: 10.1002/wcs.1232
- Carpendale, J. I., and Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Dev.* 67, 1686–1706. doi: 10.1111/j.1467-8624.1996.tb01821.x
- Carruthers, P. (2013). Mindreading in infancy. *Mind Lang.* 28, 141–172. doi: 10.1111/mila.12014
- Cassidy, J., Parke, R. D., Butkovsky, L., and Braungart, J. M. (1992). Family-peer connections: the roles of emotional expressiveness within the family and children’s understanding of emotions. *Child Dev.* 63, 603–618. doi: 10.2307/1131349
- Cassidy, K. W. (1998). Three- and four-year-old children’s ability to use desire- and belief-based reasoning. *Cognition* 66, B1–B11. doi: 10.1016/S0010-0277(98)00008-0
- Castelli, F. (2006). The Valley task: Understanding intention from goal-directed motion in typical development and autism. *Br. J. Dev. Psychol.* 24, 655–668. doi: 10.1348/026151005X54209
- Cermolacce, M., Lazerges, P., Da Fonseca, D., Fakra, E., Adida, M., Belzeaux, R., et al. (2011). Théorie de l’esprit et schizophrénie. [Theory of mind and schizophrenia.]. *L’Encephale Rev. Psychiatr. Clin. Biol. Ther.* 37, S117–S122. doi: 10.1016/S0013-7006(11)70037-9
- Chandler, M. J., and Helm, D. (1984). Developmental changes in the contribution of shared experience to social role-taking competence. *Int. J. Behav. Dev.* 7, 145–156. doi: 10.1016/S0163-6383(84)80207-6
- Chen, M. J., and Lin, Z. X. (1994). Chinese preschoolers’ difficulty with theory-of-mind tests. *Bull. Hong Kong Psychol. Soc.* 32–33, 34–46.
- Choi, Y. J., and Luo, Y. (2015). 13-month-olds’ understanding of social interactions. *Psychol. Sci.* 26, 274–283. doi: 10.1177/0956797614562452
- Chung, Y. S., Barch, D., and Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophr. Bull.* 40, 602–616. doi: 10.1093/schbul/sbt048
- Colonnesi, C., Rieffe, C., Koops, W., and Perucchini, P. (2008). Precursors of a theory of mind: a longitudinal study. *Br. J. Dev. Psychol.* 26, 561–577. doi: 10.1348/026151008X285660
- Comte-Gervais, I., Giron, A., Soares-Boucaud, I., and Poussin, G. (2008). Assessment of social intelligence in children with specific language impairment: presentation of an assessing scale. *L’Evol. Psychiatr.* 73, 353–366. doi: 10.1016/j.evopsy.2008.02.004
- Cornish, K., Rinehart, N., Gray, K., and Howlin, P. (2010). *Comic Strip Task*. Melbourne, VIC: Monash University Developmental Neuroscience and Genetic Disorders Laboratory and Monash University Centre for Developmental Psychiatry and Psychology.
- Crowe, L., Beauchamp, M., Catroppa, C., and Anderson, V. (2011). Social function assessment tools for children and adolescents: a systematic review from 1988 to 2010. *Clin. Psychol. Rev.* 31, 767–785. doi: 10.1016/j.cpr.2011.03.008
- Cutting, A. L., and Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: individual differences and interrelations. *Child Dev.* 70, 853–865. doi: 10.1111/1467-8624.00061
- Davis, P. E., Meins, E., and Fernyhough, C. (2011). Self-knowledge in childhood: relations with children’s imaginary companions and understanding of mind. *Br. J. Dev. Psychol.* 29, 680–686. doi: 10.1111/j.2044-835X.2011.02038.x
- Davis, T. L. (1998). “Children’s understanding of false beliefs in different domains: affective versus physical,” in *Poster Presented at the Conference on Human Development* (Mobile, AL).
- de Villiers, J., and de Villiers, P. (2000). “Linguistic determinism and the understanding of false beliefs,” in *Children’s Reasoning and the Mind*, eds P. Mitchell and K. J. Riggs (Hove: Psychology Press), 191–228.
- Denham, S. A. (1986). Social cognition, prosocial behavior, and emotion in preschoolers: contextual validation. *Child Dev.* 57, 194–201. doi: 10.2307/1130651
- Denham, S. A., Zoller, D., and Couchoud, E. A. (1994). Socialization of preschoolers’ emotion understanding. *Dev. Psychol.* 30, 928–936. doi: 10.1037/0012-1649.30.6.928
- Dennis, M., Simic, N., Bigler, E. D., Abildskov, T., Agostino, A., Taylor, H., et al. (2013). Cognitive, affective, and conative theory of mind (TOM) in children with traumatic brain injury. *Dev. Cogn. Neurosci.* 5, 25–39. doi: 10.1016/j.dcn.2012.11.006
- Dennis, M., Simic, N., Taylor, H., Bigler, E. D., Rubin, K., Vannatta, K., et al. (2012). Theory of mind in children with traumatic brain injury. *J. Int. Neuropsychol. Soc.* 18, 908–916. doi: 10.1017/S1355617712000756
- Dore, R. A., and Lillard, A. S. (2015). Theory of mind and children’s engagement in fantasy worlds. *Imagin. Cogn. Pers.* 34, 230–242. doi: 10.1177/0276236614568631
- Dörrenberg, S., Rakoczy, H., and Liszkowski, U. (2018). How (not) to measure infant theory of mind: testing the replicability and validity of four non-verbal measures. *Cogn. Dev.* 46, 12–30. doi: 10.1016/j.cogdev.2018.01.001
- Ebersbach, M., Stiehler, S., and Asmus, P. (2011). On the relationship between children’s perspective taking in complex scenes and their spatial drawing ability. *Br. J. Dev. Psychol.* 29, 455–474. doi: 10.1348/026151010X504942
- Eddy, C. M., and Cavanna, A. E. (2013). Altered social cognition in Tourette syndrome: nature and implications. *Behav. Neurol.* 27, 15–22. doi: 10.1155/2013/417516
- Edelstein, W., Keller, M., and Wahlen, K. (1984). Structure and content in social cognition: conceptual and empirical analyses. *Child Dev.* 55, 1514–1526. doi: 10.2307/1130021
- Feshbach, N. D., and Cohen, S. (1988). Training affect comprehension in young children: an experimental evaluation. *J. Appl. Dev. Psychol.* 9, 201–210. doi: 10.1016/0193-3973(88)90023-8
- Filippova, E., and Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Dev.* 79, 126–138. doi: 10.1111/j.1467-8624.2007.01115.x
- Fink, E., Begeer, S., Peterson, C. C., Slaughter, V., and de Rosnay, M. (2015). Friends, friendlessness, and the social consequences of gaining a theory of mind. *Br. J. Dev. Psychol.* 33, 27–30. doi: 10.1111/bjdp.12080
- Flavell, J. H. (1968). *The Development of Role-Taking and Communication Skills in Children*. New York, NY: Wiley.

- Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children's knowledge about visual perception: further evidence for the level 1–level 2 distinction. *Dev. Psychol.* 17, 99–103. doi: 10.1037/0012-1649.17.1.99
- Flavell, J. H., Green, F. L., Flavell, E. R., Watson, M. W., and Campione, J. C. (1986). Development of knowledge about the appearance-reality distinction. *Monogr. Soc. Res. Child Dev.* 51, 1–87. doi: 10.2307/1165866
- Ford, R. M., Driscoll, T., Shum, D., and Macaulay, C. E. (2012). Executive and theory-of-mind contributions to event-based prospective memory in children: exploring the self-projection hypothesis. *J. Exp. Child Psychol.* 111, 468–489. doi: 10.1016/j.jecp.2011.10.006
- Frick, A., Mohring, W., and Newcombe, N. S. (2014). Picturing perspectives: development of perspective-taking abilities in 4- to 8-year-olds. *Front. Psychol.* 5:386. doi: 10.3389/fpsyg.2014.00386
- Frith, C. D., and Frith, U. (2006). The neural basis of mentalizing. *Neuron* 50, 531–534. doi: 10.1016/j.neuron.2006.05.001
- Frith, U., Happe, F., and Siddons, F. (1994). Autism and theory of mind in everyday life. *Soc. Dev.* 3, 108–124. doi: 10.1111/j.1467-9507.1994.tb00031.x
- Galende, N., de Miguel, M. S., and Arranz, E. (2011). The role of physical context, verbal skills, non-parental care, social support, and type of parental discipline in the development of TOM capacity in five-year-old children. *Soc. Dev.* 20, 845–861. doi: 10.1111/j.1467-9507.2011.00625.x
- Gallagher, H. L., and Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends Cogn. Sci.* 7, 77–83. doi: 10.1016/S1364-6613(02)00025-6
- Garner, P. W., Curenton, S. M., and Taylor, K. (2005). Predictors of mental state understanding in preschoolers of varying socioeconomic backgrounds. *Int. J. Behav. Dev.* 29, 271–281. doi: 10.1177/01650250544000053
- Garner, P. W., Jones, D. C., and Miner, J. L. (1994). Social competence among low-income preschoolers: emotion socialization practices and social cognitive correlates. *Child Dev.* 65, 622–637. doi: 10.2307/1131405
- Geangu, E. (2002). Affirmation and negation. *Cogn. Creier Comp.* 6, 253–282. Available online at: <http://www.cbbjournal.ro/index.php/en/2002/34-6-3/177-affirmation-and-negation>
- German, T. C., and Cohen, A. S. (2012). A cue-based approach to 'theory of mind': re-examining the notion of automaticity. *Br. J. Dev. Psychol.* 30, 45–58. doi: 10.1111/j.2044-835X.2011.02055.x
- Gliga, T., Senju, A., Pettinato, M., Charman, T., and Johnson, M. H. (2014). Spontaneous belief attribution in younger siblings of children on the autism spectrum. *Dev. Psychol.* 50, 903–913. doi: 10.1037/a0034146
- Gordis, E. W., Rosen, A. B., and Grand, S. (1989). "Young children's understanding of simultaneous conflicting emotions," in *Paper Presented at the Biennial Meeting of the Society for Research in Child Development* (Kansas City, MO).
- Gratch, G. (1964). Response alternation in children: a developmental study of orientations to uncertainty. *Vita Humana* 7, 49–60. doi: 10.1159/000270053
- Guajardo, N. R., Petersen, R., and Marshall, T. R. (2013). The roles of explanation and feedback in false belief understanding: a microgenetic analysis. *J. Genet. Psychol.* 174, 225–252. doi: 10.1080/00221325.2012.682101
- Hadwin, J., Baron-Cohen, S., Howlin, P., and Hill, K. (1997). Does teaching theory of mind have an effect on the ability to develop conversation in children with autism? *J. Autism Dev. Disord.* 27, 519–537. doi: 10.1023/A:1025826009731
- Happé, F., Cook, J. L., and Bird, G. (2017). The structure of social cognition: in(ter)dependence of sociocognitive processes. *Annu. Rev. Psychol.* 68, 243–267. doi: 10.1146/annurev-psych-010416-044046
- Happé, F., and Frith, U. (2014). Annual research review: towards a developmental neuroscience of atypical social cognition. *J. Child Psychol. Psychiatry Allied Discipl.* 55, 553–557. doi: 10.1111/jcpp.12162
- Happé, F. G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord.* 24, 129–154. doi: 10.1007/BF02172093
- Harris, P. L., Donnelly, K., Guz, G. R., and Pitt-Watson, R. (1986). Children's understanding of the distinction between real and apparent emotion. *Child Dev.* 57, 895–909. doi: 10.2307/1130366
- Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., and Cooke, T. (1989). Young children's theory of mind and emotion. *Cogn. Emot.* 3, 379–400. doi: 10.1080/02699938908412713
- Hayward, E. O., and Homer, B. D. (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *Br. J. Dev. Psychol.* 35, 454–462. doi: 10.1111/bjdp.12186
- Healey, K. M., Bartholomeusz, C. F., and Penn, D. L. (2016). Deficits in social cognition in first episode psychosis: a review of the literature. *Clin. Psychol. Rev.* 50, 108–137. doi: 10.1016/j.cpr.2016.10.001
- Hedger, J. A., and Fabricius, W. V. (2011). True belief belies false belief: recent findings of competence in infants and limitations in 5-year-olds, and implications for theory of mind development. *Rev. Philos. Psychol.* 2, 429–447. doi: 10.1007/s13164-011-0069-9
- Henry, J. D., Cowan, D. G., Lee, T., and Sachdev, P. S. (2015). Recent trends in testing social cognition. *Curr. Opin. Psychiatry* 28, 133–140. doi: 10.1097/YCO.0000000000000139
- Henry, J. D., von Hippel, W., Molenberghs, P., Lee, T., and Sachdev, P. S. (2016). Clinical assessment of social cognitive function in neurological disorders. *Nat. Rev. Neurol.* 12, 28–39. doi: 10.1038/nrneuro.2015.229
- Heyes, C. (2014). False belief in infancy: a fresh look. *Dev. Sci.* 17, 647–659. doi: 10.1111/desc.12148
- Hiller, R. M., Weber, N., and Young, R. L. (2014). The validity and scalability of the theory of mind scale with toddlers and preschoolers. *Psychol. Assess.* 26, 1388–1393. doi: 10.1037/a0038320
- Hirai, M., Muramatsu, Y., Mizuno, S., Kurahashi, N., Kurahashi, H., and Nakamura, M. (2013). Developmental changes in mental rotation ability and visual perspective-taking in children and adults with Williams syndrome. *Front. Hum. Neurosci.* 7:856. doi: 10.3389/fnhum.2013.00856
- Hoddenbach, E., Koot, H. M., Clifford, P., Gevers, C., Clauser, C., Boer, F., et al. (2012). Individual differences in the efficacy of a short theory of mind intervention for children with autism spectrum disorder: a randomized controlled trial. *Trials* 13:206. doi: 10.1186/1745-6215-13-206
- Hogrefe, G.-J., Wimmer, H., and Perner, J. (1986). Ignorance versus false belief: a developmental lag in attribution of epistemic states. *Child Dev.* 57, 567–582. doi: 10.2307/1130337
- Hoogenhout, M., and Malcolm-Smith, S. (2014). Theory of mind in autism spectrum disorder: does DSM classification predict development? *Res. Autism Spectr. Disord.* 8, 597–607. doi: 10.1016/j.rasd.2014.02.005
- Houssa, M., Mazzone, S., and Nader-Grosbois, N. (2014). Validation of a French version of the theory of mind inventory (ToMI-vf). *Eur. Rev. Appl. Psychol.* 64, 169–179. doi: 10.1016/j.erap.2014.02.002
- Howlin, P., Baron-Cohen, S., and Hadwin, J. (1999). *Teaching Children With Autism to Mind-Read: A Practical Guide*. Chichester: Wiley.
- Hughes, C., Adlam, A., Happe, F., Jackson, J., Taylor, A., and Caspi, A. (2000). Good test–retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *J. Child Psychol. Psychiatry Allied Discipl.* 41, 483–490. doi: 10.1111/1469-7610.00633
- Hughes, C., Soares-Boucaud, I., Hochmann, J., and Frith, U. (1997). Social behaviour in pervasive developmental disorders: effects of informant, group and "theory-of-mind". *Eur. Child Adolesc. Psychiatry* 6, 191–198. doi: 10.1007/BF00539925
- Hutchins, T. L., Bonazinga, L. A., Prelock, P. A., and Taylor, R. S. (2008a). Beyond false beliefs: the development and psychometric evaluation of the perceptions of children's theory of mind measure-experimental version (PCToMM-E). *J. Autism Dev. Disord.* 38, 143–155. doi: 10.1007/s10803-007-0377-1
- Hutchins, T. L., Prelock, P. A., and Bonazinga, L. (2012). Psychometric evaluation of the Theory of Mind Inventory (ToMI): a study of typically developing children and children with autism spectrum disorder. *J. Autism Dev. Disord.* 42, 327–341. doi: 10.1007/s10803-011-1244-7
- Hutchins, T. L., Prelock, P. A., and Chace, W. (2008b). Test-retest reliability of a theory of mind task battery for children with autism spectrum disorders. *Focus Autism Other Dev. Disabl.* 23, 195–206. doi: 10.1177/1088357608322998
- Iannotti, R. J. (1978). Effect of role-taking experiences on role taking, empathy, altruism, and aggression. *Dev. Psychol.* 14, 119–124. doi: 10.1037/0012-1649.14.2.119
- Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., and Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: a meta-analytic review. *Dev. Psychol.* 52, 1192–1205. doi: 10.1037/dev0000140
- Jin, X., Li, P., He, J., and Shen, M. (2017). Cooperation, but not competition, improves 4-year-old children's reasoning about others' diverse desires. *J. Exp. Child Psychol.* 157, 81–94. doi: 10.1016/j.jecp.2016.12.010
- Killen, M., Lynn Mulvey, K., Richardson, C., Jampol, N., and Woodward, A. (2011). The accidental transgressor: morally-relevant theory of mind. *Cognition* 119, 197–215. doi: 10.1016/j.cognition.2011.01.006

- Kim, Y.-S., and Phillips, B. (2014). Cognitive correlates of listening comprehension. *Read. Res. Q.* 49, 269–281. doi: 10.1002/rrq.74
- Kimhi, Y. (2014). Theory of mind abilities and deficits in autism spectrum disorders. *Top. Lang. Disord.* 34, 329–343. doi: 10.1097/TLD.0000000000000033
- Knafo, A., Zahn-Waxler, C., Davidov, M., Van Hulle, C., Robinson, J. L., and Rhee, S. H. (2009). Empathy in early childhood: genetic, environmental, and affective contributions. *Ann. N. Y. Acad. Sci.* 1167, 103–114. doi: 10.1111/j.1749-6632.2009.04540.x
- Korkman, M., Kirk, U., and Kemp, S. (2007). *Clinical and Interpretive Manual NEPSY-II*. San Antonio, TX: Harcourt Assessment.
- Krcmar, M., and Vieira, E. T. Jr. (2005). Imitating life, imitating television: the effects of family and television models on children's moral reasoning. *Commun. Res.* 32, 267–294. doi: 10.1177/0093650205275381
- Kristen, S., Sodian, B., Thoermer, C., and Perst, H. (2011). Infants' joint attention skills predict toddlers' emerging mental state language. *Dev. Psychol.* 47, 1207–1219. doi: 10.1037/a0024808
- Kulke, L., von Duhn, B., Schneider, D., and Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychol. Sci.* 29, 888–900. doi: 10.1177/0956797617747090
- Lackner, C., Sabbagh, M. A., Hallinan, E., Liu, X., and Holden, J. J. (2012). Dopamine receptor D4 gene variation predicts preschoolers' developing theory of mind. *Dev. Sci.* 15, 272–280. doi: 10.1111/j.1467-7687.2011.01124.x
- Laranjo, J., Bernier, A., Meins, E., and Carlson, S. M. (2010). Early manifestations of children's theory of mind: the roles of maternal mind-mindedness and infant security of attachment. *Infancy* 15, 300–323. doi: 10.1111/j.1532-7078.2009.00014.x
- Lecce, S., Bianco, F., Devine, R. T., Hughes, C., and Banerjee, R. (2014). Promoting theory of mind during middle childhood: a training program. *J. Exp. Child Psychol.* 126, 52–67. doi: 10.1016/j.jecp.2014.03.002
- Leekam, S. (2016). Social cognitive impairment and autism: what are we trying to explain? *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150082. doi: 10.1098/rstb.2015.0082
- Leslie, A. M. (1987). Pretense and representation: the origins of "theory of mind". *Psychol. Rev.* 94, 412–426. doi: 10.1037//0033-295X.94.4.412
- Loukusa, S., Mäkinen, L., Kuusikko-Gauffin, S., Ebeling, H., and Moilanen, I. (2014). Theory of mind and emotion recognition skills in children with specific language impairment, autism spectrum disorder and typical development: group differences and connection to knowledge of grammatical morphology, word-finding abilities and verbal working memory. *Int. J. Lang. Commun. Disord.* 49, 498–507. doi: 10.1111/1460-6984.12091
- Luke, N., and Banerjee, R. (2013). Differentiated associations between childhood maltreatment experiences and social understanding: a meta-analysis and systematic review. *Dev. Rev.* 33, 1–28. doi: 10.1016/j.dr.2012.10.001
- Martin, A. K., Robinson, G., Dzafic, I., Reutens, D., and Mowry, B. (2014). Theory of mind and the social brain: implications for understanding the genetic basis of schizophrenia. *Genes Brain Behav.* 13, 104–117. doi: 10.1111/gbb.12066
- Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., and Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Dev.* 45, 357–366. doi: 10.2307/1127956
- Mayes, L. C., Klin, A., Tercyak, K. P. Jr., Cicchetti, D. V., and Cohen, D. J. (1996). Test-retest reliability for false-belief tasks. *J. Child Psychol. Psychiatry Allied Discipl.* 37, 313–319. doi: 10.1111/j.1469-7610.1996.tb01408.x
- McDonald, S. (2013). Impairments in social cognition following severe traumatic brain injury. *J. Int. Neuropsychol. Soc.* 19, 231–246. doi: 10.1017/S1355617712001506
- McKown, C., Russo-Ponsaran, N. M., Johnson, J. K., Russo, J., and Allen, A. (2016). Web-based assessment of children's social-emotional comprehension. *J. Psychoeduc. Assess.* 34, 322–338. doi: 10.1177/0734282915604564
- Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Dev. Psychol.* 31, 838–850. doi: 10.1037/0012-1649.31.5.838
- Meltzoff, A. N., and Brooks, R. (2008). Self-experience as a mechanism for learning about others: a training study in social cognition. *Dev. Psychol.* 44, 1257–1265. doi: 10.1037/a0012888
- Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychol. Bull.* 135, 749–773. doi: 10.1037/a0016854
- Mitchell, P., Saltmarsh, R., and Russell, H. (1997). Overly literal interpretations of speech in autism: understanding that messages arise from minds. *J. Child Psychol. Psychiatry Allied Discipl.* 38, 685–691. doi: 10.1111/j.1469-7610.1997.tb01695.x
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* 4:1. doi: 10.1186/2046-4053-4-1
- Moll, H., Kane, S., and McGowan, L. (2016). Three-year-olds express suspense when an agent approaches a scene with a false belief. *Dev. Sci.* 19, 208–220. doi: 10.1111/desc.12310
- Moll, H., Koring, C., Carpenter, M., and Tomasello, M. (2006). Infants determine others' focus of attention by pragmatics and exclusion. *J. Cogn. Dev.* 7, 411–430. doi: 10.1207/s15327647jcd0703\_9
- Moll, H., and Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *Brit. J. Dev. Psychol.* 24, 603–613. doi: 10.1348/026151005X55370
- Moll, H., and Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Dev. Psychol.* 43, 309–317. doi: 10.1037/0012-1649.43.2.309
- Muris, P., Steerneman, P., Meesters, C., Merckelbach, H., Horselenberg, R., van den Hogen, T., et al. (1999). The TOM test: a new instrument for assessing theory of mind in normal children and children with pervasive developmental disorders. *J. Autism Dev. Disord.* 29, 67–80. doi: 10.1023/A:1025922717020
- Nader-Grosbois, N., Houssa, M., and Mazzone, S. (2013). How could theory of mind contribute to the differentiation of social adjustment profiles of children with externalizing behavior disorders and children with intellectual disabilities? *Res. Dev. Disabil.* 34, 2642–2660. doi: 10.1016/j.ridd.2013.05.010
- Onishi, K. H., and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* 308, 255–258. doi: 10.1126/science.1107621
- Oppenheimer, L., and Rempt, E. (1986). Social cognitive development with moderately and severely retarded children. *J. Appl. Dev. Psychol.* 7, 237–249. doi: 10.1016/0193-3973%2886%2990032-8
- O'Reilly, K., Peterson, C. C., and Wellman, H. M. (2014). Sarcasm and advanced theory of mind understanding in children and adults with prelingual deafness. *Dev. Psychol.* 50, 1862–1877. doi: 10.1037/a0036654
- Payne, J. M., Porter, M., Pride, N. A., and North, K. N. (2016). Theory of mind in children with neurofibromatosis type 1. *Neuropsychology* 30, 439–448. doi: 10.1037/neu0000262
- Perner, J., Leekam, S. R., and Wimmer, H. (1987). Three-year-olds' difficulty with false belief: the case for a conceptual deficit. *Br. J. Dev. Psychol.* 5, 125–137. doi: 10.1111/j.2044-835X.1987.tb01048.x
- Perner, J., and Wimmer, H. (1985). "John thinks that Mary thinks that...": attribution of second-order beliefs by 5- to 10-year-old children. *J. Exp. Child Psychol.* 39, 437–471. doi: 10.1016/0022-0965(85)90051-7
- Peskin, J., and Astington, J. (2004). The effects of adding metacognitive language to story texts. *Cogn. Dev.* 19, 253–273. doi: 10.1016/j.cogdev.2004.01.003
- Peskin, J., Prusky, C., and Comay, J. (2014). Keeping the reader's mind in mind: development of perspective-taking in children's dictations. *J. Appl. Dev. Psychol.* 35, 35–43. doi: 10.1016/j.appdev.2013.11.001
- Peterson, C. C., Garnett, M., Kelly, A., and Attwood, T. (2009). Everyday social and conversation applications of theory-of-mind understanding by children with autism-spectrum disorders or typical development. *Eur. Child Adolesc. Psychiatry* 18, 105–115. doi: 10.1007/s00787-008-0711-y
- Phillips, A. T., and Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition* 98, 137–155. doi: 10.1016/j.cognition.2004.11.005
- Phillips, A. T., Wellman, H. M., and Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition* 85, 53–78. doi: 10.1016/S0010-0277(02)00073-2
- Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *J. Exp. Child Psychol.* 47, 116–129. doi: 10.1016/0022-0965(89)90066-0
- Poletti, M., and Adenzalo, M. (2013). Theory of mind in non-autistic psychiatric disorders of childhood and adolescence. *Clin. Neuropsychiatry* 10, 188–195. Available online at: [https://www.researchgate.net/publication/259177200\\_Theory\\_of\\_mind\\_in\\_non-autistic\\_psychiatric\\_disorders\\_of\\_childhood\\_and\\_adolescence](https://www.researchgate.net/publication/259177200_Theory_of_mind_in_non-autistic_psychiatric_disorders_of_childhood_and_adolescence)
- Pons, F., and Harris, P. L. (2000). *Test of Emotion Comprehension-TEC*. Oxford: University of Oxford.

- Poulin-Dubois, D., Sodian, B., Metz, U., Tilden, J., and Schoeppner, B. (2007). Out of sight is not out of mind: developmental changes in infants' understanding of visual perception during the second year. *J. Cogn. Dev.* 8, 401–425. doi: 10.1080/15248370701612951
- Poulin-Dubois, D., and Yott, J. (2014). Executive functions and theory of mind understanding in young children: a reciprocal relation? *Psychol. Franc.* 59, 59–69. doi: 10.1016/j.psfr.2013.11.002
- Powell, L. J., Hobbs, K., Bardis, A., Carey, S., and Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cogn. Dev.* 46, 40–50. doi: 10.1016/j.cogdev.2017.10.004
- Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/S0140525X00076512
- Reed, T. (1994). Performance of autistic and control subjects on three cognitive perspective-taking tasks. *J. Autism Dev. Disord.* 24, 53–66. doi: 10.1007/BF02172212
- Repacholi, B. M., and Gopnik, A. (1997). Early reasoning about desires: evidence from 14- and 18-month-olds. *Dev. Psychol.* 33, 12–21. doi: 10.1037/0012-1649.33.1.12
- Ribordy, S. C., Camras, L. A., Stefani, R., and Spaccarelli, S. (1988). Vignettes for emotion recognition research and affective therapy with children. *J. Clin. Child Psychol.* 17, 322–325. doi: 10.1207/s15374424jccp1704\_4
- Ricard, M., Girouard, P. C., and Decarie, T. G. (1999). Personal pronouns and perspective taking in toddlers. *J. Child Lang.* 26, 681–697. doi: 10.1017/s0305000999003943
- Rieffe, C., Terwogt, M. M., Koops, W., Stegge, H., and Oomen, A. (2001). Preschoolers' appreciation of uncommon desires and subsequent emotions. *Brit. J. Dev. Psychol.* 19, 259–274. doi: 10.1348/026151001166065
- Riggio, M. M., and Cassidy, K. W. (2009). Preschoolers' processing of false beliefs within the context of picture book reading. *Early Educ. Dev.* 20, 992–1015. doi: 10.1080/10409280903375685
- Ruffman, T., and Olson, D. (1989). Children's ascriptions of knowledge to others. *Dev. Psychol.* 25, 601–606. doi: 10.1037/0012-1649.25.4.601
- Russell, J., Mauthner, N., Sharpe, S., and Tidswell, T. (1991). The “windows task” as a measure of strategic deception in preschoolers and autistic subjects. *Br. J. Dev. Psychology* 9, 331–349. doi: 10.1111/j.2044-835X.1991.tb00881.x
- Sabbagh, M. A., and Paulus, M. (2018). Replication studies of implicit false belief with infants and toddlers. *Cogn. Dev.* 46, 1–3. doi: 10.1016/j.cogdev.2018.07.003
- Schult, C. A. (2002). Children's understanding of the distinction between intentions and desires. *Child Dev.* 73, 1727–1747. doi: 10.1111/1467-8624.t01-1-00502
- Scott, R. M., and Baillargeon, R. (2017). Early false-belief understanding. *Trends Cogn. Sci.* 21, 237–249. doi: 10.1016/j.tics.2017.01.012
- Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *Neuroscientist* 18, 108–113. doi: 10.1177/1073858410397208
- Senju, A., Southgate, V., Snape, C., Leonard, M., and Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychol. Sci.* 22, 878–880. doi: 10.1177/0956797611411584
- Shaked, M., and Yirmiya, N. (2004). Matching procedures in autism research: evidence from meta-analytic studies. *J. Autism Dev. Disord.* 34, 35–40. doi: 10.1023/B:JADD.0000018072.42845.83
- Slaughter, V. (2015). Theory of mind in infants and young children: a review. *Aust. Psychol.* 50, 169–172. doi: 10.1111/ap.12080
- Slaughter, V., Imuta, K., Peterson, C., and Henry, J. D. (2015). Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Dev.* 86, 1159–1174. doi: 10.1111/cdev.12372
- Slick, D. J. (2006). “Psychometrics in neuropsychological assessment,” in *A Compendium of Neuropsychological Tests*, eds E. Strauss, E. M. S. Sherman, and O. Spreen (Oxford: Oxford University Press), 3–32.
- Smiley, P. A. (2001). Intention understanding and partner-sensitive behaviors in young children's peer interactions. *Soc. Dev.* 10, 330–354. doi: 10.1111/1467-9507.00169
- Smith, N. (1973). *Empathy and the Egocentric Child*. Ames, IA: Iowa State University.
- Smogorzewska, J., Szumski, G., and Grygiel, P. (2019). The children's social understanding scale: an advanced analysis of a parent-report measure for assessing theory of mind in Polish children with and without disabilities. *Dev. Psychol.* 55, 835–845. doi: 10.1037/dev0000673
- Snodgrass, C., and Knott, F. (2006). Theory of mind in children with traumatic brain injury. *Brain Injury* 20, 825–833. doi: 10.1080/02699050600832585
- Sodian, B., Licata, M., Kristen-Antonow, S., Paulus, M., Killen, M., and Woodward, A. (2016). Understanding of goals, beliefs, and desires predicts morally relevant theory of mind: a longitudinal investigation. *Child Dev.* 87, 1221–1232. doi: 10.1111/cdev.12533
- Song, M. J., Choi, H. I., Jang, S.-K., Lee, S.-H., Ikezawa, S., and Choi, K.-H. (2015). Theory of mind in Koreans with schizophrenia: a meta-analysis. *Psychiatry Res.* 229, 420–425. doi: 10.1016/j.psychres.2015.05.108
- Southgate, V., Senju, A., and Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychol. Sci.* 18, 587–592. doi: 10.1111/j.1467-9280.2007.01944.x
- Sprong, M., Schothorst, P., Vos, E., Hox, J., and van Engeland, H. (2007). Theory of mind in schizophrenia: meta-analysis. *Br. J. Psychiatry* 191, 5–13. doi: 10.1192/bjp.bp.107.035899
- Sprung, M. (2010). Clinically relevant measures of children's theory of mind and knowledge about thinking: non-standard and advanced measures. *Child Adolesc. Ment. Health* 15, 204–216. doi: 10.1111/j.1475-3588.2010.00568.x
- Stanzione, C., and Schick, B. (2014). Environmental language factors in theory of mind development: evidence from children who are deaf/hard-of-hearing or who have specific language impairment. *Top. Lang. Disord.* 34, 296–312. doi: 10.1097/TLD.0000000000000038
- Steerneman, P., Jackson, S., Pelzer, H., and Muris, P. (1996). Children with social handicaps: an intervention programme using a theory of mind approach. *Clin. Child Psychol. Psychiatry* 1, 251–563. doi: 10.1177/1359104596012006
- Steinmann, E., Schmalor, A., Prehn-Kristensen, A., Wolff, S., Galka, A., Mohring, J., et al. (2014). Developmental changes of neuronal networks associated with strategic social decision-making. *Neuropsychologia* 56, 37–46. doi: 10.1016/j.neuropsychologia.2013.12.025
- Stewart, E., Catroppa, C., and Lah, S. (2016). Theory of mind in patients with epilepsy: a systematic review and meta-analysis. *Neuropsychol. Rev.* 26, 3–24. doi: 10.1007/s11065-015-9313-x
- Strasser, K., and del Rio, F. (2014). The role of comprehension monitoring, theory of mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. *Read. Res. Q.* 49, 169–187. doi: 10.1002/rrq.68
- Sullivan, K., Winner, E., and Hopfield, N. (1995). How children tell a lie from a joke: the role of second-order mental state attributions. *Br. J. Dev. Psychol.* 13, 191–204. doi: 10.1111/j.2044-835X.1995.tb00673.x
- Suway, J. G., Degnan, K. A., Sussman, A. L., and Fox, N. A. (2012). The relations among theory of mind, behavioral inhibition, and peer interactions in early childhood. *Soc. Dev.* 21, 331–342. doi: 10.1111/j.1467-9507.2011.00634.x
- Swettenham, J. (1996). Can children be taught to understand false belief using computers? *Child Psychol. Psychiatry Allied Discipl.* 37, 157–165. doi: 10.1111/j.1469-7610.1996.tb01387.x
- Tager-Flusberg, H., and Sullivan, K. (1994). Predicting and explaining behavior: a comparison of autistic, mentally retarded and normal children. *J. Child Psychol. Psychiatry* 35, 1059–1075. doi: 10.1111/j.1469-7610.1994.tb01809.x
- Tager-Flusberg, H., and Sullivan, K. (2000). A componential view of theory of mind: evidence from Williams syndrome. *Cognition* 76, 59–90. doi: 10.1016/S0010-0277(00)00069-X
- Tahiroglu, D., Moses, L. J., Carlson, S. M., Mahy, C. E., Olofson, E. L., and Sabbagh, M. A. (2014). The children's social understanding scale: construction and validation of a parent-report measure for assessing individual differences in children's theories of mind. *Dev. Psychol.* 50, 2485–2497. doi: 10.1037/a0037914
- Tardif, T., Wellman, H. M., and Cheung, K. M. (2004). False belief understanding in Cantonese-speaking children. *J. Child Lang.* 31, 779–800. doi: 10.1017/S0305000904006580
- Tashakkori, A., and Teddlie, C. (Eds.). (2010). *SAGE Handbook of Mixed Methods in Social & Behavioral Research. 2nd Edn.* Thousand Oaks, CA: SAGE.
- Taylor, M. (1988). Conceptual perspective taking: children's ability to distinguish what they know from what they see. *Child Dev.* 59, 703–718. doi: 10.2307/1130570
- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* 60, 34–42. doi: 10.1016/j.jclinepi.2006.03.012

- Turkstra, L. S., Abbeduto, L., and Meulenbroek, P. (2014). Social cognition in adolescent girls with fragile X syndrome. *Am. J. Intellect. Dev. Disabil.* 119, 319–339. doi: 10.1352/1944-7558-119.4.319
- Veneziano, E., and Christian, H. (2006). *Etats Internes, Fausse Croyance et Explications Dans les Récits: Effets de l'étayage chez les Enfants de 4 à 12 ans.* (Langage et l'Homme), 117–138.
- Vera-Estay, E., Dooley, J.J., and Beauchamp, M.H. (2015). Cognitive underpinnings of moral reasoning in adolescence: the contribution of executive functions. *J. Moral Educ.* 44, 17–33. doi: 10.1080/03057240.2014.986077
- Viranyi, Z., Topal, J., Miklosi, A., and Csanyi, V. (2006). A nonverbal test of knowledge attribution: a comparative study on dogs and children. *Anim. Cogn.* 9, 13–26. doi: 10.1007/s10071-005-0257-z
- Vuadens, P. (2005). The anatomical bases of the theory of mind: a review. *Schweiz. Arch. Neurol. Psychiatr.* 156, 136–146. doi: 10.4414/sanp.2005.01600
- Walz, N. C., Yeates, K. O., Taylor, H., Stancin, T., and Wade, S. L. (2010). Theory of mind skills 1 year after traumatic brain injury in 6- to 8-year-old children. *J. Neuropsychol.* 4, 181–195. doi: 10.1348/174866410X488788
- Wang, Y., Liu, H., and Su, Y. (2014). Development of preschoolers' emotion and false belief understanding: a longitudinal study. *Soc. Behav. Pers.* 42, 645–654. doi: 10.2224/sbp.2014.42.4.645
- Wellman, H. A., Cross, D., and Watson, J. (2001). Meta-analysis of theory of mind development: the truth about false belief. *Child Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304
- Wellman, H. M., and Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition* 30, 239–277. doi: 10.1016/0010-0277(88)90021-2
- Wellman, H. M., Fang, F., and Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: longitudinal perspectives. *Child Dev.* 82, 780–792. doi: 10.1111/j.1467-8624.2011.01583.x
- Wellman, H. M., and Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Dev.* 75, 523–541. doi: 10.1111/j.1467-8624.2004.00691.x
- Wellman, H. M., and Woolley, J. D. (1990). From simple desires to ordinary beliefs: the early development of everyday psychology. *Cognition* 35, 245–275. doi: 10.1016/0010-0277(90)90024-E
- Westby, C. E. (2014). Social neuroscience and theory of mind. *Folia Phoniatr. Logoped.* 66, 7–17. doi: 10.1159/000362877
- Whitehouse, A., and Hird, K. (2004). Is grammatical competence a precondition for belief-desire reasoning? Evidence from typically developing children and those with autism. *Adv. Speech Lang. Pathol.* 6, 39–51. doi: 10.1080/14417040410001669480
- Williamson, R. A., Brooks, R., and Meltzoff, A. N. (2015). The sound of social cognition: Toddlers' understanding of how sound influences others. *J. Cogn. Dev.* 16, 252–260. doi: 10.1080/15248372.2013.824884
- Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5
- Yagmurlu, B., Berument, S. K., and Celimli, S. (2005). The role of institution and home contexts in theory of mind development. *J. Appl. Dev. Psychol.* 26, 521–537. doi: 10.1016/j.appdev.2005.06.004
- Yirmiya, N., Erel, O., Shaked, M., and Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychol. Bull.* 124, 283–307. doi: 10.1037/0033-2909.124.3.283
- Ziatabar Ahmadi, S. Z., Jalaie, S., and Ashayeri, H. (2015). Validity and reliability of published comprehensive theory of mind tests for normal preschool children: a systematic review. *Iran. J. Psychiatry* 10, 214–224.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Beaudoin, Leblanc, Gagner and Beauchamp. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.