The relationship between Theory of Mind and Executive Functioning in typically developing children: A systematic review

Dimitra Economacou, Alexandra Karousou, Nikolaos Makris

doi: 10.12681/psy_hps.33349

Copyright © 2023, Dimitra Economacou, Alexandra Karousou, Nikolaos K. Makris

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0.

To cite this article:

The relationship between Theory of Mind and Executive Functioning in typically developing children: A systematic review

Dimitra ECONOMACOU¹, Alexandra KAROUSOU², Nikolaos MAKRIS¹

¹ Department of Primary Education, Democritus University of Thrace
² Department of Education Sciences in Early Childhood, Democritus University of Thrace

KEYWORDS
Theory of Mind (ToM), mental state understanding, executive functions, cognitive development, typical development, children

ABSTRACT
The development of Theory of Mind (ToM) constitutes a central topic in the study of human cognitive development. During the last decades, researchers representing different theoretical approaches have been trying to explore the emergence and development of mental state understanding and the potential contribution of other complex cognitive mechanisms in this developmental process. The relationship between ToM and Executive Functions (EFs) has been assessed in the context of various experimental designs. Despite the significant number of existing studies and the variety of relevant theories proposed, there is still no agreement about the exact nature of the ToM-EFs relationship. The main aim of this paper is to review relevant research results on typically developing children. It attempts to disentangle the factors that can potentially explain the contradictory findings reported in the literature. The results, overall, support the ToM-EFs relationship and suggest EFs' important role in ToM development. However, the exact nature of this relationship seems obscured by the diversity of approaches, operationalization of the theoretical constructs, methods and ages included in the studies. It is, therefore, suggested that for building a unified picture and an explanatory account of the dynamic developmental relationship between these two complex theoretical constructs, a refinement in the conceptual definitions and methodological approaches is crucial.

CORRESPONDENCE
Dimitra Economacou,
Democritus University of Thrace, Department of Primary Education,
Nea Chili, 68100 Alexandroupolis, Greece.
dimoikonomakou@gmail.com

Introduction
In the general research area on the development of children's knowledge about the mental world, Theory of Mind constitutes one of the most prominent domains of theoretical and empirical inquiry. The theoretical construct of Theory of Mind (ToM) encompasses the awareness of the representational character of the other's mind or the human mind in general (Flavell, 2004). Recently, a vast body of research on ToM development has pointed to its potential relationship with Executive Functions (EFs). Primarily reported for Autism Spectrum Disorder populations (Russel et al., 1991), findings concerning the similar developmental trajectory of ToM and EFs, as well as regarding their joint impairment, sparked a scientific debate on their relationship and became a hot topic in neurodevelopmental psychology. Despite the great number of studies, both on typically and atypically developing children and adults, the exact nature and extent of ToM and EFs interconnection remain controversial. The lack of a unifying conceptual framework for both ToM and EFs, the use of a great variety of tasks to assess their various components, and the varying ages of the samples included in relevant studies are only some of the factors that could potentially underlie this controversy.

Given the inherent theoretical interest of the topic, as well as its importance for understanding children's cognitive development, the present paper aims to systematically review the existing research results concerning the relationship between ToM and EFs in an attempt to disentangle the factors that underlie -and can potentially explain- the contradictory findings reported in the literature.
Theory of Mind: conceptualisation and development

ToM, referred to in the literature also as ‘mind-reading’, ‘mentalizing’, or ‘mental state attribution’ (Kemp et al., 2012), is the capacity to attribute desires, knowledge, intentions, beliefs, and emotions to the self and to others. To do so, a person must construct different abstract representations of reality and navigate between them while distinguishing between their own mental representations and those of others (Wimmer & Perner, 1983). Given its implications for an individual’s social and cognitive functioning, ToM has become a pervasive research topic in various disciplines, such as social psychology, cognitive psychology, neuropsychology, developmental psychology, social neuroscience, or speech therapy.

This multidisciplinary inquiry into ToM has yielded numerous operationalisations of the theoretical construct. Depending on their particular research interests and ages of inquiry, researchers from different scientific backgrounds focused on different aspects of ToM development, linking different sub-categories of abilities to the theoretical construct of ToM. In their systematic review, Beaudoin et al. (2020) summarized seven (7) basic categories and 39 sub-categories of abilities linked by different researchers to ToM: (a) intentions (intention explanations, completion of failed actions, discrepant intentions, prediction of actions, intention attribution to visual figures), (b) beliefs (location false beliefs, identity false beliefs, second-order false beliefs, content false beliefs, belief based actions/emotions, sequence false beliefs), (c) desires (discrepant desires, desire-action contradiction, multiple desires, desires influences on emotions and actions), (d) emotions (moral emotions, typical emotional reactions, atypical emotional reactions, mixed emotions, emotion regulation, discrepant emotions, hidden emotions), (e) knowledge (knowledge-attention links, percepts-knowledge links, information-knowledge links, knowledge-pretend play links), (f) mentalistic understanding of non-literal communication (white lies, irony/sarcasm, egocentric lies, faux pas, humor, involuntary lies), and (g) percepts (auditory perspective taking, percept-action link, complex visual perspective taking, simple visual perspective taking). As it becomes apparent, ToM is an umbrella term encompassing a wide range of abilities and concepts, ranging from simple desire understanding to more complex social behaviours (e.g., lying, irony and humour).

Overall, the timeline of “ToM development” seems to vary among studies, depending on the exact aspect of ToM that each researcher -explicitly or implicitly- includes under the umbrella concept of ToM and the respective methodological tools they chose for its evaluation (German & Cohen, 2012). It is generally accepted that infants seem to understand desires, intentions, and the causal relation between a person’s emotions and goals as early as around 14-18 months of age (Yott & Poulin-Dubois, 2016). The explicit attribution of False Belief (FB), recognized as a cognitively demanding capacity, is not achieved until 3 to 5 years of age (Wellman et al., 2001). Pre-schoolers have also been shown to distinguish their beliefs and knowledge of the world from others (Brüne & Brüne-Cohrs, 2006). Five to 6-year-old children can realize that someone may hold beliefs about another person’s beliefs (Wimmer & Perner, 1983). Finally, more complex social behaviours like using metaphors, irony, jokes, and the capacity to judge socially inappropriate behaviours appear to emerge in middle childhood and early adolescence (Brüne & Brüne-Cohrs, 2006).

However, the above plurality in the conceptualisations of ToM and the resulting disparity in the measurement tools (in theory, all measuring ToM, but in reality assessing different asynchronously developing aspects of it) has indirectly obscured the attempts to investigate how ToM relates to other cognitive abilities, as is the case discussed in this paper. As will be seen later, research looking into associations between ToM and EFs usually and (often implicitly) addresses associations between one or two concrete aspects of ToM and one or two aspects of EFs in different ages, making comparisons and generalisation of the findings a challenging task.

Executive Functions: conceptualisation and development

Executive functions (EFs) have been widely approached as a critical factor for ToM development for at least the past 20 years (Dahlgren et al., 2017; Hughes, 1998). However, the term EFs is also an umbrella concept, encompassing a variety of cognitive processes: planning, working memory, self-monitoring, self-regulation, initiation, inhibition, and attention (Goldstein et al., 2014). In general, the term EFs refers to all those cognitive processes that support complex, goal-directed behaviour, particularly in situations that are novel or require conscious effort (Diamond, 2006; Zelazo et al., 2003). Among the most well-known and widely accepted models, Miyake et al. (2000) propose that there are three main EF aspects: a) shifting (SH) between mental states or
tasks, b) updating and monitoring of Working Memory (WM) representations, and c) inhibition of dominant and prepotent responses, also known as inhibitory control (IC).

It is not surprising, therefore, that there is not a single timeline of “EFs development”. Shifting, also referred to as mental flexibility, gradually reaches the level of adult performance by age 12 (Crone et al., 2009). Working memory, the system responsible for actively processing and maintaining information for a short time, follows a linear pattern of age-related improvement from 4 to 15 years (Gathercole et al., 2004). Self-regulation, which according to Luria (1959), constitutes a prerequisite for inhibition, emerges between 3 and 5 years of age, while inhibition, the cognitive ability responsible for suppressing a dominant response (Diamond, 2006), exhibits a significant improvement between 5 and 11 years of age (Romine & Reynolds, 2005). In 2004, Kerr and Zelazo proposed a further distinction of the broad notion of EFs by introducing the terms hot and cool EFs. While cool EFs are responsible for problem-solving in decontextualized situations, hot executive functioning gets involved when motivation and emotions can influence a behavioural decision (Zelazo et al., 2005). Despite the small number of research assessing the development of hot EFs, children seem to be able to successfully regulate their behaviour, considering an upcoming reward by the age of 5 years (Bunch & Andrews, 2012).

Overall, depending again on the exact aspect of EF being researched and the methodological tools used for its evaluation, the timeline of EF development may differ among studies, covering an extended period from infancy to young adulthood (Wilson et al., 2018). Using an umbrella term, sometimes without an explicit definition and informed selection of its aspects chosen to address or without a precise interpretation of its findings, can undermine the attempts to investigate how EFs relate to other concurrently developing cognitive abilities. As mentioned earlier and will be examined in detail in the results, research looking into associations between EFs and ToM most of the time (and often implicitly) addresses associations between one or two EFs and one or two concrete aspects of ToM. This fragmentary approach makes comparisons and generalisation of the findings challenging.

The relationship between ToM and EF: theoretical background

During the last decades, researchers representing different theoretical and scientific backgrounds have tried to pinpoint the cognitive mechanisms that may be related to and affect ToM development (Devine et al., 2016). Studies evaluating the cognitive profile of non-typically developing children, principally of individuals with Autism Spectrum Disorder (ASD), offered perhaps the first indications of an EFs - ToM relationship (Baron-Cohen et al., 1999). Russell et al. (1991) and Hughes and Russell (1993) first noted that improvements in executive performance might contribute to subsequent ToM development. Subsequently published data further highlighted the existence of significant correlations between ToM and EFs (Carlson & Moses, 2001; Oh & Lewis, 2008; Sabbagh, Moses et al., 2006). An association between EFs and ToM, or, more specifically, False Belief understanding (FBU), was also detected in meta-analytic works (Devine et al., 2016; Perner and Lang, 1999, 2000). However, negative results concerning the EF - ToM relationship are also reported in the literature (e.g., Dahlgren et al., 2017).

In parallel with these developments, several contemporary studies aimed to dig into the nature of this relationship and propose a theoretical explanation. Two principal accounts have been proposed to explain ToM and EFs interconnection. The “expression” account (Moses, 2003), on the one hand, suggests that children understand mental states but, for some reason, fail to manifest this understanding. For example, it is claimed that in an unexpected location task, where an object is moved in the absence of the task's main character, children may fail to report the character's beliefs about the object's location because they find it hard to suppress their own knowledge concerning the object's actual location (Carlson et al., 2014). This account received support from several studies showing that decreasing a task's executive demands could improve performance in pre-schoolers (Leslie & Polizzi, 1998). On the other hand, the “emergence” account (Moses, 2003) suggests that EFs are necessary for acquiring ToM concepts. According to this approach, setting aside one's point of view requires inhibitory control and is considered a prerequisite for simultaneously taking into account distinct perspectives. Longitudinal studies showing that earlier EFs performance predicts later ToM skills have supported the “emergence theory” (Hughes & Ensor, 2005; Müller et al., 2012).

Again, the lack of a consensus regarding the theoretical basis of a potential EFs-ToM relationship could be attributed to the complex nature of both theoretical constructs. As mentioned earlier, the concrete dimensions
of ToM and EFs that each researcher chooses to include in their study may significantly affect the results. The underlying conceptualisation of the terms may set barriers to the generalisability of results, even indirectly, as it is directly linked with the tasks selected to measure each construct. In the same vein, the wide variety of measures developed for assessing ToM and EFs do not necessarily target the same abilities, nor are they all psychometrically validated (German & Cohen, 2012). Finally, the varying ages of the participants combined with the differing conceptualisations of ToM and EFs may lead to completely different results given the different developmental trajectories for the various dimensions included in them.

**Aim of the systematic review**

Based on the above rationale, the picture of the ToM - EFs relationship often seems partial or fragmentary. In our view, important questions remain unanswered in the literature. For instance, when one claims that ToM is related to EFs, for which of their dimensions is this claim valid? All of them? Some of them? In all ages? In some ages? As far as we know, there is still no clear answer.

The main aim of this paper is to review research results on the nature of the ToM - EFs relationship in typically developing children, considering and accounting for the important diversity of approaches, conceptualisations, methods and ages included in the studies reviewed. We believe this is a necessary first step towards obtaining an unfragmented picture of the relationship between the two constructs, as well as a prerequisite step for gaining a clearer understanding of the cognitive and developmental processes that underlie this relationship.

**Method**

Data reported in this study were extracted in November 2022, and all studies published by that time were included in the sample. Three research strategies were used to include a representative sample of studies in the current systematic review. Key electronic databases (Scopus, Pubmed, EBSCO and Google Scholar) were systematically searched for papers assessing the relationship between ToM and EFs. In addition, the "Cognitive Development Society" members were asked to participate in this research by sharing their relevant experimental results. Finally, every review and book chapter retrieved from databases was further examined as a new source of possibly relevant articles.

Aiming to include all papers studying the interconnection between ToM and EFs in typically developing populations, ToM and EFs have been included in this search using multiple equivalent terms. More specifically, the following key search terms have been searched in titles, keywords and abstracts: (1) ToM, (2) EFs, (3) children, and (4) typical development were adequately combined, as follows: (1) "Theory of Mind" OR "mental state attribution" OR "mind-reading"), (2) "executive functions" OR "inhibition" OR "mental flexibility" OR "working memory" OR "planning"), (3) "children". The combination of the three research strategies resulted in a total of 4725 articles.

Exclusion criteria were adopted to ensure the inclusion of all empirical papers reporting on the relationship between EFs and ToM in typically developing children while excluding results on atypical development or other extraneous factors, as well as to ensure the methodological rigour of the studies included in the review. Therefore, studies focusing on neurodevelopmental disorders, neurological pathologies, sensory disabilities, and cognitive impairment were excluded unless they had a control group and separately reported results on EFs and ToM relationship in the typically developing controls. Therefore, 42 studies not having a control group in their experimental design were excluded (e.g., Kenny et al., 2019; Tager-Flusberg et al., 2005; Yu et al., 2021). In addition, the family context was taken into consideration, given the existing data showing a relationship between secure attachment and ToM performance (Arranz et al., 2002). For this reason, three studies, including institutionalized children, were eliminated as not representative of the general population (Colvert et al., 2008; Etel & Yagmurlu, 2014; Selcuk & Yucel, 2017). One study was excluded due to the methodological weaknesses reported in the sample, which could potentially affect the results (Hughes et al., 1998). A prevalent issue in the papers retrieved was that several studies did not report on the relationship between EFs and ToM directly. Thus, in many papers, the principal research question was focused either on the mutual impact of EFs and ToM on other social or cognitive skills (e.g., sharing, general cognitive development) or on the contribution of some other...
cognitive mechanisms (e.g., phonological awareness) on EFs and ToM performance. Among those cases, 45 papers, after being carefully evaluated, were excluded for reporting results exclusively on the relationship of ToM with the third involved variable and not providing data on the relationship between ToM and EFs (e.g., Demetriou et al., 2021; Matthews et al., 2018; Liu et al., 2016; Venkadasalam et al., 2022; Wang et al., 2022).

Next, to ensure the empirical quality of the papers included in the review, the use of neuropsychological/experimental tests for assessing both EFs and ToM was considered a prerequisite for inclusion. Despite presenting a moderate level of ecological validity (Chaytor & Schmitter-Edgecombe, 2003), neuropsychological/experimental tests are considered to offer more reliable results. Therefore, eight studies following a different approach for evaluating sample performance (e.g., assessment via mobile application tests, Klindt et al., 2017) were excluded. Another quality criterion for inclusion consisted of adopting a minimum sample size of fifty (50) participants since it is considered the minimum sample size to guarantee a reasonable power in a within-participants design (Brysbaert, 2019). Therefore, 114 studies were excluded as their sample size was lower than fifty (e.g., Dahlgren et al., 2017; Farrar & Ashwell, 2012; Kloo et al., 2010). Additionally, two case studies (Markiewicz et al., 2009; Goukon et al., 2006) and two papers published in a non-English language (Karakele & Ertuglur, 2012; Thommen et al., 2016) were also excluded. Finally, 87 non-empirical publications (book chapters, reviews) were also excluded.

Table 1 contains a brief presentation of the inclusion and exclusion criteria set before the beginning of our research. Figure 1 constitutes a PRISMA Flow Chart describing the process followed and the number of eliminated papers in every selection phase.

The screening and classification process was performed independently by two authors. The reliability of the decisions in both cases was excellent (91% and 93.5%, respectively). The papers for which there was a disagreement between the two independent evaluators were decided by mutual consensus.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typically developing children</td>
<td>Non-typically developing children</td>
</tr>
<tr>
<td>Absence of variables that could bias the results</td>
<td>Presence of variables that could bias the results</td>
</tr>
<tr>
<td>(family context, individual characteristics)</td>
<td>(family context, individual characteristics)</td>
</tr>
<tr>
<td>Empirical studies (longitudinal, cross-sectional,</td>
<td>Case studies, book chapters, reviews, systematic</td>
</tr>
<tr>
<td>training programs/protocols)</td>
<td>reviews</td>
</tr>
<tr>
<td>Studies using neuropsychological tests/experimental</td>
<td>Studies not using neuropsychological tests/experimental</td>
</tr>
<tr>
<td>tasks</td>
<td></td>
</tr>
<tr>
<td>Number of participants ≥ 50</td>
<td>Number of participants &lt;50</td>
</tr>
<tr>
<td>Presence of statistical indices concerning the</td>
<td>Absence of statistical indices concerning the</td>
</tr>
<tr>
<td>relationship between EFs and ToM</td>
<td>relationship between EFs and ToM</td>
</tr>
<tr>
<td>Papers published in English</td>
<td>Papers published in other languages</td>
</tr>
</tbody>
</table>

For a total of 111 papers that were finally included, the following data have been gathered: sample size, age range of participants, country, number and duration of sessions in training programs, interval between different assessment time points in longitudinal studies, tasks used for the assessment of EFs, and finally, results. These data were used to classify the studies and interpret their results and will be further discussed in the following section.
Records identified through database searching (n=4725)

Additional Records identified through other sources (n=16) (book, reviews, meta-analysis)

Results after removal of duplicates (n=4601)

Records excluded based on their titles (n=3734)

Abstracts retrieved after removal of non-relevant articles based on screening title of study (n=867)

Records excluded based on their abstract (n=589)

Full-text articles retrieved (n=278)

Articles included in the qualitative synthesis (n=111)

Full-text articles excluded (N=166), with respect to the following exclusion criteria:
- Environmental/Individual characteristics (n=4)
- Experimental design (n=55): (case study: n=2) & (non-experimental tests: n=8) & (not providing statistical proof for correlation: n=45)
- Sample < 50 participants (control group): n=18,
- Foreign language (n=2)
- Non empirical publications (books, reviews, systematic-review) (n=87)

Figure 1. PRISMA flow chart describing the levels of screening and the process of selection of included studies
Results

Data classification

The analysis of 111 studies led to their primary organisation into three categories depending on their experimental design. Cross-sectional (n=87), longitudinal (n=18), and training protocols (n=6) are presented separately in Tables 2, 3, 4 and 5. Furthermore, given the prevalent number of studies focusing uniquely on False Belief (FB) understanding as a ToM measure, cross-sectional studies are presented into two groups: studies including exclusively FB tasks (Table 2) and studies assessing multiple aspects of ToM (e.g., attribution of desires, emotion recognition) (Table 3). In the FB category, experimenters mostly used stories presenting the unexpected location, the unexpected content or the difference between appearance and reality paradigms to evaluate children’s false belief attribution. It should be noted that Wimmer and Perner’s unexpected location test (1983), Flavell et al.’s (1983, 1986) appearance reality task, and the Smarties test by Gopnik & Astington (1988) were among the most frequently used FB tests (see, Table 2). When multiple facets of ToM skills were assessed, the batteries most commonly used were: Strange Stories (Happé et al., 1994), ToM Scale (Wellman & Liu, 2004), and NEPSY II Social Perception tasks (Korkman et al., 2007) (see, Table 3). When the experimental design did not include an already existing battery, several individual tasks were administered to assess attribution of desires, perception, pretence, knowledge awareness, deception and emotion recognition/ attribution. It should be noted that FB understanding was evaluated almost in all the studies included in the review.

Given the differing developmental timelines of both the ToM and EF subdomains, the second primary criterion for the organization of the studies was the participants’ chronological age. As a result, five broad age-range categories were formed: infancy (0 to 18 months), toddlerhood (18 to 36 months), early childhood (36 to 74 months), middle childhood (74-144 months), and adolescence (144-168 months) (Craig & Baucum, 2002). As seen in Table 2, the relationship between FB understanding and EFs is predominately assessed in early and middle childhood. On the contrary, studies using a variety of ToM tests recruited samples of a broader age range (Table 3). This difference in age groups is in line with the existing data showing that FB understanding develops by the age of 6 years. In addition, the complex nature of the tasks explicitly assessing ToM and EFs leads to limited studies on infants and toddlers. Finally, training programs were primarily applied in early and middle childhood (see Table 4).

In order to account for the dimensions of ToM and EFs that researchers evaluated in their study, we opted for coding the cognitive abilities addressed in each paper rather than the name of the tests used. FB tests were reported as Unexpected Location (UL), Unexpected Content (UC), and Appearance-Reality (AR) tests, according to the event that led the story’s protagonist to have a FB. The rest of ToM measures were reported according to the type of mental state that was being attributed (Deception (DC), Pretence (PR), Emotion Understanding (EU), Knowledge Access (KA), Social Stories / Faux pas (SS), Different Desires (DD), Intentions Understanding (IU), Perspective Taking (PT)). This way of categorisation was not always feasible since some tests required the simultaneous attribution of different types of mental states (intentions, perspectives, beliefs). This is why they are referred to as Mental State Attribution Tasks (MSA).

We should note that ToM batteries, such as the ToM scale (Wellman & Liu, 2004), NEPSY-II ToM tasks (Korkman et al., 2007), and Strange Stories (Happé et al., 1994), are coded in Table 3 with their full name for convenience reasons. However, in the results section, they are analysed according to the various dimensions of ToM they assess. Classifying tests according to the skill assessed was more difficult for EFs measures. The multilevel cognitive demands of EFs tasks, and the frequent lack of information concerning the reasons for selecting each measure, complicated this effort. Unfortunately, many of the most well-known tasks of EF assessment, for example, the DCCS (Zelazo, 2006), were found to be used for the assessment of different EFs, depending on the experimental design of each study (SH in Cantin et al., 2016, IC in Kloé et al., 2010; Chu & Minal, 2018; Scullin & Bonner, 2006, assessment of reasoning in Müller et al., 2005). WM and IC were the two most frequently assessed EFs. For the evaluation of WM, the Digit Span task (mostly backward) (Davis & Pratt, 1995) and the Counting and Labelling tasks were among the most broadly used tests (Gordon & Olson, 1998). On the other hand, IC was either assessed with gratification / hot IC (ICH) tasks (Diamond, 2013), such as the Gift Delay (Kochanska et al., 1996) or the Snack Delay task (Kochanska et al., 2000) or with cool IC tasks (ICC),
as Go/NoGo-like (Carlson et Moses, 2001) and Stroop-like tasks (e.g., Bear/Dragon: Reed et al., 1984; Day/Night: Gerstadt et al., 1994).

**Relationship between FB and EFs**

Only six of the studies included failed to confirm a relationship between EFs and FB despite the use of multiple tests for the assessment of EFs (Cardillo et al., 2021; Diaz & Farrar, 2018b; Meinhardt-Injac et al., 2020; Scullin & Bonner, 2006; Sudo & Matsui, 2021 -only for the monolingual sample; Talwar et al., 2017).

Conversely, a relationship between FB and EFs was evidenced in all the remaining studies assessing FB. More specifically, significant correlations have been found in children aged 3 – 9 years old between cool IC and performance in FB tasks, either as the only evaluated EF (e.g., Bellagamba et al., 2015; Dick et al., 2005; Lang & Perner, 2002; Poole et al., 2014) or as part of a broader EF battery (e.g., Grosse Wiesmann et al., 2017; McGlamery et al., 2007). Moreover, cool IC has also been found to predict FB score in children aged 3-12 (Cassetta et al., 2018; Kouklari et al., 2017; Powell & Carey, 2017; Wang et al., 2012). As for the relationship between hot IC with FB, three studies failed to find a link between the delay of gratification tasks and FB performance (Bellagamba et al., 2015; Carlson et al., 2002; Chasiotis et al., 2006). The remaining studies detected either a correlation (Grosse Wiesmann et al., 2017; Sabbagh, Moses et al., 2006) between hot IC and FB (3-5.5 years) or an effect of hot IC on FB performance (4-10 years) (Miller et al., 2018; Powell & Carey, 2017).

Working Memory has been identified as a significant predictor of children’s ability to correctly attribute beliefs to others in the unexpected transfer or unexpected content tasks. The impact of WM on FB performance was the principal research question for many studies, which were either uniquely focused on this cognitive mechanism (Keenan, 1998) or compared its contribution to ToM to that of other EFs (Buac & Kaushanskaya, 2020-only for the monolingual sample; Kennedy et al., 2015; Longobardi et al., 2021; Mutter et al., 2006). In addition, performance in WM tasks was correlated to FB performance (ages 3-12 years) (Gordon et al., 2014; Hasselhorn et al., 2005; Talwar et Lee, 2008), a result which remained significant even when the participants’ justifications of their responses in the mental states’ attribution task were taken into account (Ford et al., 2012; Putko & Zlotogorska, 2014).

Finally, FB performance has been associated with other aspects of EFs, such as SH (ages 3-8 years) (Blair & Razza, 2007; Gordon et al., 2014; Guajardo et al., 2009; Kara & Selcuk, 2021; Oh & Lewis, 2008) and planning (PL) (Putko & Zlotogorska, 2014). Furthermore, children’s scores (ages 8-11 years) in SH (Cassetta et al., 2018) and PL (ages 3-6 years) (Cole & Mitchell, 2000; Longobardi et al., 2021; Memisevic et al., 2018) have also been found to predict FB performance.

**Attribution of desires and EFs**

Attribution of desires is one of the fundamental aspects of ToM. The ability to understand that someone else does not necessarily share the same desires with us or that their actions do not always reflect their desires has been explored even in infancy. These are usually measured by tasks assessing the understanding and attribution of conflicting desires (e.g., Desire Appreciation Task or Conflicting Desires Task). Indeed, attribution of desires is found to be one of the ToM abilities with the earliest onset (Peterson & Wellman, 2019) and is considered less cognitively demanding than FB. Only two studies have assessed the association between attribution of desires and EFs by means of a dedicated task. In particular, a significant correlation has been found between a conflicting desires task and IC in preschoolers (ages 3-5 years) (Rakoczy, 2010). Also, performance in the DCCS task has been found to be a predictor of conflicting desires (ages 4-7 years) (Rostad & Pexman, 2014).

**Attribution of intentions and EFs**

Attribution of intentions, however, despite being central to human perspective-taking and interpersonal relations (Maselli & Altrocchi, 1969), has not been widely addressed in the EFs-ToM literature. Attribution of intentions is usually assessed with picture sequencing tasks. Only one study has addressed the effect of cool and hot IC on attribution of intentions in toddlers and has found positive results (Carlson et al., 2004).
Knowledge Access - Perspective Taking and EFs

Explaining a person’s behaviour by evaluating their possible knowledge of a particular situation is essential to understanding and responding to their social behaviours. Knowledge access has been mainly assessed in preschoolers, for example, through the “Hiding Games” task, and is found to be predicted by their performance in tasks assessing cool IC (Chu & Minai, 2018).

Highly linked with knowledge access, perspective-taking is an additional key function for social understanding. More specifically, when trying to explain the behaviour of others, one should acknowledge that each person perceives the world differently and that these different perceptions (auditory, visual, etc.) lead one to act in particular ways. The ability to acknowledge and adopt someone else’s perception of the world activates complex mental functions, such as the visualisation and the mental rotation of visual stimuli, in the case of visual perception. This ability is taken to express complex cognitive mechanisms and is found to be predicted by SH performance (ages 6-10 years) (Miller et al., 2018). However, according to Carlson, Mandell and Williams (2004) age seems to play an important role in this relationship. In the second wave of their longitudinal study, IC (hot and cool) was correlated to visual perspective taking in 39 months, whereas no relationship was found when the same participants were 24 months old.

Pretence and Deception and EFs

Pretend play has always presented a great interest for researchers of ToM development. Pretence seems to share some underlying skills with mental state understanding. According to Lillard (1993), pretence is a mental state that requires not only a mental representation of an object but also the capacity to inhibit reality to adopt an alternative representation. The findings of the present review revealed that pretence was correlated to cool and hot IC, WM and SH (ages 2-3 years) (Carlson, Mandell et al., 2004; Hughes & Ensor, 2006). Importantly, two studies found it could both predict and be predicted by IC, SH and PL (ages 2-6 years) (Hughes & Ensor, 2005; McAlister & Peterson, 2006).

Similarly, deception constitutes another ability highly related to mental state understanding. The ability to infer the other person’s mental representation is considered a prerequisite for deception, even when this knowledge remains implicit. For example, in their study, Chandler and his colleagues (1998) revealed that 2 to 3 years old children erased a puppet’s footprints, willing to keep the location of their treasure secret. The performance in deceptive tasks, even of different levels of complexity, has been correlated to IC, WM, and SH (Guajardo et al., 2009; Hoyo et al., 2019; Hughes & Ensor, 2006) (sample age range: 3-9 years old). Furthermore, an effect of IC on deception in preschoolers (3-5 years old) (Chasiotis et al., 2006), but also a bidirectional effect between IC, SH and deception (Hughes & Ensor, 2005) was reported in toddlers (2-3 years old).

Understanding of emotions and EFs

Understanding another person’s emotional state is commonly found in the ToM literature as part of the affective ToM (Dennis et al., 2013). Some of the basic forms of relevant ToM tasks require understanding mixed, discrepant, or hidden emotions, inferring their causes, and estimating typical vs atypical emotional reactions. Emotion understanding (EU) was found to be correlated with WM in middle childhood (ages 6-11 years) (Morra et al., 2011) and SH (ages 5-9 years old) (Hoyo et al., 2019). Additionally, exploring the relationship between EU and IC in similar age groups (5-8 years old), Hudson and Jacques (2014) found only the recognition of discrepant emotions to be related to IC. Importantly, a series of studies in a wide age range of participants (3.5-18 years old) has confirmed an effect of IC, ICH, SH, WM or PL (Draperi et al., 2022; Gabriel et al., 2019; Kouklari et al. 2017; Longobardi, 2021; Vetter et al., 2013) on emotion understanding. Emotion understanding will be further discussed in the context of ToM scales presentation in the next section.

ToM Scales and EFs

Children’s performance on ToM scales and its relationship with executive functioning is presented separately due to the complex nature of these assessment tools. Wellman and Liu’s ToM scale (2004) is considered one of the most complete batteries, offering a multilevel assessment of ToM (current version assesses
knowledge access, different desires, different beliefs, contents false belief, explicit false belief, belief emotion, hidden emotion; Peterson & Wellman, 2019). Used in different cultural contexts and various age groups, the ToM Scale has contributed to a more comprehensive study of ToM development. IC (cool and hot), SH, PL, and WM are the main EFs that have been correlated either to the total score of the ToM Scale tasks (ages 3-12) (Baker et al., 2021; Fujita et al., 2022; Karpinski & Scullin, 2009; Melinder et al., 2006; Nathanson & Fries, 2014; Wade et al., 2014) or to particular subtests (Doenyas et al., 2018; FB, EU and knowledge access subscales). Moreover, WM (ages 3-11 years) (Burnel et al., 2020; Duh et al., 2016; Shahaeian et al., 2015), IC (sample's age range: 4-5) (Duh et al., 2016; Shahaeian et al., 2015), and SH (ages 4-5 years) (Shahaeian et al., 2015) were found to predict the ToM Scale's total score. Finally, Henning and colleagues (2011) were able to confirm the effect of the DCCS score on the ToM Scale's total score, but they could not confirm the inverse effect (ages 3-6 years).

The Strange Stories test (Happé et al., 1994), initially created for assessing autistic children, includes 24 social stories. The main purpose of this test is to examine a person's ability to understand non-literal communication concepts, such as irony, white lies, sarcasm, and humour (Beaudoin et al., 2020). It should be noted that the Strange Stories task was used only for assessing older children (approximately 6-17 years old). Significant correlations were detected between the Strange Stories total score and WM, IC cool and hot, and SH (ages 6-12 years) (Austin et al., 2014; Bock et al., 2015; Wilson et al., 2021). Performance in the Strange Stories task was also found to be predicted by IC and SH in adolescence (ages 13-18 years) (Gabriel et al., 2019) and also by WM (ages 6-18 years old) (Cantin et al., 2016; Gabriel et al., 2019). Im-Bolter and colleagues (2016) have shown that according to the participant’s age, the contribution of EFs may differ. More specifically, 7-12 years old children's scores in Strange Stories were found to be positively predicted by WM, SH, and IC, while adolescents’ performance by SH and IC.

Finally, the NEPSY-II test, and more specifically, its Social Perception subscale, assesses the following skills: facial affect recognition, comprehension of others' perspectives, intentions and beliefs. Social perception subtests of NEPSY-II were used in a small number of studies as a measure of mental states' understanding in different social contexts. WM, IC and PL were found to be correlated to the children’s (3-16 years-old) performance in the NEPSY-II ToM subtests (Brock et al., 2019; Huyder et al., 2017), although Huyder et al. (2017) failed to find a correlation with PL for the 9-12 years group. Finally, a significant effect of WM and the EF scale of NEPSI-II was found on the NEPSY-II ToM subtests (ages 3-6 years) (Rosenqvist et al., 2014).

Training and longitudinal studies

Training programs and longitudinal studies do not only provide evidence about the relationship between EF and ToM. A deeper analysis of their results could offer a much deeper insight into the nature of this relationship and provide evidence of potential causalities. This section will further assess the developmental interconnection between ToM and EFs based on the results of both training programs and longitudinal studies.

Despite their limited number, the training programs present inconclusive results on the predictive role of EFs in ToM performance. Of the six reported studies (Table 4), one focused on training both EFs and ToM skills (Kloo & Perner, 2003). By including two different experimental groups, each exposed to either EFs or ToM tasks, Kloo and Perner revealed a bidirectional developmental link between EFs and ToM. Three of the remaining training studies, exclusively focusing on ToM training, identified pre-training EFs as a significant predictor of post-training ToM performance (Gao et al., 2020; Lecce & Bianco, 2018; Qu et al., 2015). Finally, two training programs did not confirm the effect of EFs on ToM (Arslan et al., 2018; Lecce et al., 2014). In particular, Arslan and her colleagues (2018) noted that the difficulty of pre-training EFs assessment tasks could explain the negative findings since the performance in an easy pre-training WM test was unrelated to post-training FB performance. In the study of Lecce and colleagues (2014), participation in the training program has improved participants' performance in both EF and ToM post-training tasks. However, the observed gains in EFs performance were unrelated to post-training ToM.

Longitudinal studies included different ages from 18 months to 10 years, with intervals varying from 4 months to 4 years and differed in the EFs and the aspects of ToM they assessed. These studies offer some insight into the EF-ToM interconnected developmental trajectories, although their results are not very clear. Eleven studies found an effect of one or more EFs on some aspect of ToM at a subsequent time point (Austin et al., 2014;
**Table 2**
The relationship between False Belief Understanding and Executive Functions

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Authors¹</th>
<th>N</th>
<th>Age (in month)²</th>
<th>Country</th>
<th>ToM³</th>
<th>EF⁴</th>
<th>Relationship⁵</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD</td>
<td>Kara &amp; Sekuc, 2021</td>
<td>115</td>
<td>33-95</td>
<td>Turkey</td>
<td>UC &amp; UL (1st and 2nd order), AR</td>
<td>ICC, WM, SH</td>
<td>ToM→EF</td>
</tr>
<tr>
<td></td>
<td>Sudo &amp; Matsui, 2021</td>
<td>50</td>
<td>62-66</td>
<td>Japan</td>
<td>UL</td>
<td></td>
<td>ICC</td>
</tr>
<tr>
<td></td>
<td>DiCataldo &amp; Roch, 2020</td>
<td>111</td>
<td>44-75</td>
<td>Italy</td>
<td>UC</td>
<td></td>
<td>ICC + other variables→UL (bilinguals)</td>
</tr>
<tr>
<td></td>
<td>Diaz &amp; Farrar, 2018b</td>
<td>65</td>
<td>40-69</td>
<td>USA</td>
<td>UL, UC, AR</td>
<td>ICC, SH</td>
<td>WM→U</td>
</tr>
<tr>
<td></td>
<td>Memisevic et al., 2018</td>
<td>116</td>
<td>36-72</td>
<td>Bosnia Herzegovina</td>
<td>UL</td>
<td></td>
<td>SH + other variables→UL</td>
</tr>
<tr>
<td></td>
<td>O'Toole et al., 2017</td>
<td>106</td>
<td>46-80</td>
<td>UK</td>
<td>UL, UC</td>
<td>ICC, ICH, PL, WM</td>
<td>ICC &amp; PL &amp; WM→ToM</td>
</tr>
<tr>
<td></td>
<td>Talwar et al., 2017</td>
<td>160</td>
<td>47-71</td>
<td>Canada</td>
<td>UL</td>
<td></td>
<td>N.S.</td>
</tr>
<tr>
<td></td>
<td>Powell &amp; Carey, 2017 (1)</td>
<td>70</td>
<td>60-71</td>
<td>USA</td>
<td>UL</td>
<td></td>
<td>ICH</td>
</tr>
<tr>
<td></td>
<td>Powell &amp; Carey, 2017 (2)</td>
<td>102</td>
<td>48-71</td>
<td>USA</td>
<td>UL (prediction, justification)</td>
<td>ICC</td>
<td>ICC→ToM (even when EF demands were lower)</td>
</tr>
<tr>
<td></td>
<td>Grosse Wiesmann et al., 2017</td>
<td>57</td>
<td>36-48</td>
<td>Germany</td>
<td>UL &amp; UC (explicit), UL (implicit)</td>
<td>SH, ICC, ICH</td>
<td>SH &amp; ICC &amp; FB (explicit)</td>
</tr>
<tr>
<td></td>
<td>Bellagamba et al., 2015</td>
<td>101</td>
<td>35-49</td>
<td>Italy</td>
<td>UC</td>
<td></td>
<td>ICC, ICH</td>
</tr>
<tr>
<td></td>
<td>Kennedy et al., 2015</td>
<td>102</td>
<td>53-107</td>
<td>USA</td>
<td>FB</td>
<td></td>
<td>WM, ICC</td>
</tr>
<tr>
<td></td>
<td>Putro &amp; Zlotogorska, 2014</td>
<td>59</td>
<td>36-48</td>
<td>Portugal</td>
<td>UL &amp; UC (prediction, justification)</td>
<td>ICC, WM</td>
<td>WM + other variables→UL (justification)</td>
</tr>
<tr>
<td></td>
<td>Gordon et al., 2014</td>
<td>107</td>
<td>48-144</td>
<td>Canada</td>
<td>UL (1st and 2nd order)</td>
<td>BRIEF</td>
<td>PL &amp; SH &amp; WM→ToM</td>
</tr>
<tr>
<td></td>
<td>Poole et al., 2014</td>
<td>61</td>
<td>48-108</td>
<td>USA</td>
<td>UC, UL</td>
<td>ICC</td>
<td>ICC→UC &amp; WM</td>
</tr>
<tr>
<td></td>
<td>Ford et al., 2012 (1)</td>
<td>59</td>
<td>48-72</td>
<td>Australia</td>
<td>UC, UL</td>
<td>ICC, WM</td>
<td>WM &amp; ICC (not all tests)→ToM</td>
</tr>
<tr>
<td></td>
<td>Ford et al., 2012 (2)</td>
<td>50</td>
<td>49-67</td>
<td>Australia</td>
<td>FBv, FB</td>
<td>ICC, WM</td>
<td>ICC→FB &amp; FBv</td>
</tr>
<tr>
<td></td>
<td>Wang et al., 2012</td>
<td>192</td>
<td>36-48</td>
<td>China</td>
<td>UC, UL</td>
<td>ICC</td>
<td>ICC + other variables→ToM</td>
</tr>
<tr>
<td></td>
<td>Talwar &amp; Lee, 2008</td>
<td>150</td>
<td>36-96</td>
<td>USA</td>
<td>UL, UC (1st and 2nd order)</td>
<td>WM, ICC</td>
<td>WM &amp; ICC→ToM</td>
</tr>
<tr>
<td></td>
<td>Oh &amp; Lewis, 2008</td>
<td>138</td>
<td>36-62</td>
<td>Korea, UK</td>
<td>UL, UC</td>
<td>ICC, WM, SH</td>
<td>ICC &amp; SH→FB (KR)/ ICC &amp; SH &amp; FB (UK)</td>
</tr>
<tr>
<td></td>
<td>Müller et al., 2007</td>
<td>99</td>
<td>36-65</td>
<td>Canada</td>
<td>UL, UC</td>
<td>WM</td>
<td>WM + other variables→UL</td>
</tr>
<tr>
<td></td>
<td>Blair &amp; Razza, 2007</td>
<td>170</td>
<td>45-83</td>
<td>USA</td>
<td>UL, UC</td>
<td>ICC, SH</td>
<td>ICC→UC &amp; WM</td>
</tr>
<tr>
<td></td>
<td>Lewis et al., 2006</td>
<td>67</td>
<td>36-65</td>
<td>China</td>
<td>UC, UL</td>
<td>ICC, SH, WM</td>
<td>WM + other variables→ToM &gt; ICC + other variables→ToM</td>
</tr>
<tr>
<td></td>
<td>Mutter et al., 2006</td>
<td>72</td>
<td>32-64</td>
<td>USA</td>
<td>UC</td>
<td></td>
<td>ICC, WM</td>
</tr>
<tr>
<td></td>
<td>Sabbagh, Moses et al, 2006</td>
<td>60</td>
<td>37-58</td>
<td>Canada</td>
<td>UL, ULv, UCv</td>
<td>ICC, ICH</td>
<td>ICC+UL &amp; UL &amp; UCv/ ICH+UL</td>
</tr>
<tr>
<td></td>
<td>Scullin &amp; Bonner, 2006</td>
<td>62</td>
<td>36-60</td>
<td>USA</td>
<td>UC</td>
<td></td>
<td>N.S.</td>
</tr>
<tr>
<td></td>
<td>Hasselhun et al, 2005</td>
<td>126</td>
<td>38-69</td>
<td>Germany</td>
<td>FB (1st and 2nd order)</td>
<td>WM</td>
<td>WM→FB (1&lt;4.90)</td>
</tr>
<tr>
<td></td>
<td>Müller et al., 2005</td>
<td>69</td>
<td>37-65</td>
<td>Canada</td>
<td>UC, UCh</td>
<td>DCCS</td>
<td>DCCS + other variables→UC</td>
</tr>
<tr>
<td></td>
<td>Dick et al, 2005</td>
<td>107</td>
<td>36-60</td>
<td>USA</td>
<td>UL, AR</td>
<td>ICC</td>
<td>ICC→UL</td>
</tr>
<tr>
<td></td>
<td>Andrews et al, 2003 (1)</td>
<td>60</td>
<td>36-60</td>
<td>Australia</td>
<td>AR, UC, UL</td>
<td>ICC</td>
<td>ICC + other variables→AR &amp; UC</td>
</tr>
<tr>
<td></td>
<td>Andrews et al, 2003 (2)</td>
<td>60</td>
<td>36-60</td>
<td>Australia</td>
<td>AR, UC, UL</td>
<td>ICC</td>
<td>ICC + other variables→AR &amp; UC</td>
</tr>
<tr>
<td></td>
<td>Klo &amp; Perren, 2003</td>
<td>60</td>
<td>33-48</td>
<td>Austria</td>
<td>UL</td>
<td></td>
<td>DCCS=UL (only with 1 of 2 stories)</td>
</tr>
<tr>
<td></td>
<td>Carlson et al, 2002</td>
<td>47</td>
<td>40-66</td>
<td>USA</td>
<td>UL, UC, AR</td>
<td>ICH, ICC, WM</td>
<td>ICC + other variables→UL &amp; UC</td>
</tr>
<tr>
<td></td>
<td>Lang &amp; Pierres, 2002</td>
<td>57</td>
<td>37-61</td>
<td>Austria</td>
<td>UL</td>
<td></td>
<td>ICH</td>
</tr>
<tr>
<td></td>
<td>Cole &amp; Mitchell, 2000 (1)</td>
<td>121</td>
<td>36-48</td>
<td>UK</td>
<td>UL, UC</td>
<td>ICC, PL</td>
<td>ICC &amp; PL + other variables→UL &amp; UC</td>
</tr>
<tr>
<td></td>
<td>Cole &amp; Mitchell, 2000 (2)</td>
<td>71</td>
<td>47-68</td>
<td>UK</td>
<td>UL, UC</td>
<td>ICC, PL</td>
<td>ICC &amp; PL + other variables→UL &amp; UC</td>
</tr>
<tr>
<td></td>
<td>Keenan, 1998</td>
<td>60</td>
<td>48-62</td>
<td>New Zealand</td>
<td>UL</td>
<td></td>
<td>WM</td>
</tr>
<tr>
<td></td>
<td>Gordon &amp; Olson, 1998</td>
<td>72</td>
<td>36-76</td>
<td>Canada</td>
<td>UC, AR</td>
<td>DT</td>
<td>DT→ToM</td>
</tr>
<tr>
<td>MC</td>
<td>Buac &amp; Kaushansky, 2020</td>
<td>115</td>
<td>84-144</td>
<td>USA</td>
<td>UL (1st and 2nd order)</td>
<td>ICC, SH, WM</td>
<td>WM + other variables→UL (monolinguals)</td>
</tr>
<tr>
<td></td>
<td>Cassette et al, 2018</td>
<td>168</td>
<td>96-132</td>
<td>Canada</td>
<td>FB (2nd order) &amp; FBe</td>
<td>ICC, SH, WM</td>
<td>SH + other variables→UL (bilinguals)</td>
</tr>
<tr>
<td></td>
<td>McClanahan et al, 2007</td>
<td>66</td>
<td>69-80</td>
<td>USA</td>
<td>UC, UL (2nd order)</td>
<td>ICC, SH, WM, PL</td>
<td>EF + other variables→FB &amp; FBe</td>
</tr>
</tbody>
</table>

¹ TD = Toddlerhood, EC = Early Childhood, MC = Middle Childhood
² Numbers (1) and (2) next to the authors refer to the different experiments reported in the same paper.
³ Ages typed in bold extend beyond the limits of the age group.
⁴ ToM (Iv=low verbal): DC=Deception, PR=Pretendno; FB= False Belief, UL=Unexpected Location, MSA=Mental State Attribution, EU=Emotion Understanding, KA=Knowledge Access, AR=Appearance-Reality, TK=Think-Know (differentiation), UC=Unexpected Content, IU=Intentions understanding, PT=Perspective taking, SS=Social Stories, DD=Different Desires
⁵ EFs: WM=Working Memory, ICC=Inhibitory Control Cold, SH=shifting/switching/mental flexibility, ICH=Inhibitory Control Hot, DT=Dual Task, PL=planning, CST=card shorting test, PCA=physical causality
⁶ Symbols referring to the relationship between EFs and ToM: ↔ = variable X is related to variable Y (correlation), ↩ = variable X predicts variable Y (regression), ↔ = variable X predicts variable Y, but variable Y also predicts variable X (regression)
## Table 3

The relationship between ToM scales and Executive Functions

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Authors, Year</th>
<th>N (Age in months)</th>
<th>Country</th>
<th>ToM</th>
<th>EF</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD</td>
<td>Wade et al., 2014</td>
<td>468 2-45</td>
<td>Canada</td>
<td>ToM Scale</td>
<td>ICC, SH, WM</td>
<td>EF ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>Hughes &amp; Ensor, 2006</td>
<td>127 24-36</td>
<td>USA</td>
<td>DC, PR, FB</td>
<td>WM, ICC, SH</td>
<td>EF → ToM</td>
</tr>
<tr>
<td></td>
<td>Hughes &amp; Ensor, 2005</td>
<td>140 24-36</td>
<td>UK</td>
<td>UL, PR, DC</td>
<td>ICC, SH</td>
<td>EF ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>Draperi et al., 2022</td>
<td>112 42-78</td>
<td>France</td>
<td>FB, EU</td>
<td>ICC, ICH</td>
<td>ICH (mediating role) ↔ EU</td>
</tr>
<tr>
<td></td>
<td>Fujita et al., 2022</td>
<td>236 48-84</td>
<td>Japan &amp; UK</td>
<td>ToM Scale</td>
<td>ICC, WM, SH</td>
<td>ToM ↔ EF (both cultures)</td>
</tr>
<tr>
<td></td>
<td>Xie et al., 2006</td>
<td>109 36-60</td>
<td>China</td>
<td>UL, AR, DC</td>
<td>ICC, WM</td>
<td>EF ↔ ToM (despite better performance in EF)</td>
</tr>
<tr>
<td></td>
<td>Duh et al., 2016</td>
<td>922 36-60</td>
<td>China</td>
<td>ToM Scale</td>
<td>ICC, WM</td>
<td>ICC &amp; WM ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>Chasiotis et al., 2006</td>
<td>314 36-60</td>
<td>Germany, Costa Rica, Cameroon</td>
<td>UL, DC</td>
<td>ICC, ICH</td>
<td>ICC + other variables ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>Longobardi et al., 2021</td>
<td>144 35-71</td>
<td>Italy</td>
<td>UL, EU</td>
<td>Brief-P</td>
<td>WM + other variables ↔ UL</td>
</tr>
<tr>
<td></td>
<td>Baker et al., 2021</td>
<td>121 37-78</td>
<td>USA</td>
<td>ToM Scale</td>
<td>ICC</td>
<td>ToM ↔ ICC</td>
</tr>
<tr>
<td></td>
<td>Wilson et al., 2021</td>
<td>126 60-144</td>
<td>Australia</td>
<td>Strange stories</td>
<td>WM, ICC, ICH</td>
<td>EF ↔ ToM (only when age not controlled)</td>
</tr>
<tr>
<td></td>
<td>Austin et al., 2020</td>
<td>1657 72-132</td>
<td>Germany</td>
<td>MSA</td>
<td>ICC, SH, WM</td>
<td>WM + other variables ↔ ToM Scale</td>
</tr>
<tr>
<td></td>
<td>Burnell et al., 2020</td>
<td>126 38-141</td>
<td>France</td>
<td>ToM Scale, FBv</td>
<td>ICC, SH, WM</td>
<td>SH + other variables ↔ FBv</td>
</tr>
<tr>
<td></td>
<td>Hoyos et al., 2019</td>
<td>86 60-108</td>
<td>Spain</td>
<td>DC, UL (1st 2nd order), EU</td>
<td>WM, SH, ICC</td>
<td>SH ↔ EU (8-9 y)</td>
</tr>
<tr>
<td></td>
<td>Tsoji &amp; Mitchell, 2019</td>
<td>160 36-72</td>
<td>Japan</td>
<td>UL, IU</td>
<td>SH, ICC</td>
<td>ICC ↔ DC &amp; UL (5-6 yo)</td>
</tr>
<tr>
<td></td>
<td>Brock et al., 2019</td>
<td>140 M 2-192</td>
<td>USA</td>
<td>NEPSY II</td>
<td>WM, ICC</td>
<td>WM &amp; ICC ↔ NEPSY II</td>
</tr>
<tr>
<td></td>
<td>Miller et al., 2018</td>
<td>81 81-120</td>
<td>USA</td>
<td>PT, SS</td>
<td>SH, WM, ICH</td>
<td>SH + other variables ↔ PT</td>
</tr>
<tr>
<td></td>
<td>Chu &amp; Minal, 2018</td>
<td>354 36-75</td>
<td>USA</td>
<td>KA</td>
<td>ICC</td>
<td>ICH &amp; SH + other variables ↔ SS</td>
</tr>
<tr>
<td></td>
<td>Lagattuta et al., 2018</td>
<td>211 48-120</td>
<td>USA</td>
<td>MSA</td>
<td>ICC, WM</td>
<td>WM &amp; ICC + other variables ↔ MSA</td>
</tr>
<tr>
<td></td>
<td>Wilson et al., 2018</td>
<td>126 60-132</td>
<td>Australia</td>
<td>SS</td>
<td>ICC, ICH, WM, SH</td>
<td>EF (hot &amp; cool) ↔ SS</td>
</tr>
<tr>
<td></td>
<td>Huwyler et al., 2017</td>
<td>248 60-144</td>
<td>Switzerland, Germany</td>
<td>NEPSY II</td>
<td>PL, ICC, WM</td>
<td>PL &amp; WM ↔ NEPSY-II (younger participants)</td>
</tr>
<tr>
<td></td>
<td>Aslan et al., 2016</td>
<td>89 53-120</td>
<td>Turkey</td>
<td>UL (2nd order)</td>
<td>WM</td>
<td>WM + other variables ↔ UL (only for younger participants)</td>
</tr>
<tr>
<td></td>
<td>Shahaeian et al., 2015</td>
<td>140 48-60</td>
<td>Iran</td>
<td>ToM Scale, UL</td>
<td>ICC, WM, SH</td>
<td>EF + other variables ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>Hudson &amp; Jacques, 2014</td>
<td>107 60-96</td>
<td>Canada</td>
<td>EU</td>
<td>ICC</td>
<td>ICC ↔ EU (only for hidden emotions)</td>
</tr>
<tr>
<td></td>
<td>Lagattuta et al., 2014</td>
<td>195 48-132</td>
<td>USA</td>
<td>ToM probability task</td>
<td>ICC, WM</td>
<td>ICC + WM + other variables ↔ ToM probability task</td>
</tr>
<tr>
<td></td>
<td>Nathanson &amp; Fries, 2014</td>
<td>107 38-74</td>
<td>USA</td>
<td>ToM Scale</td>
<td>PL, ICC, WM</td>
<td>EF ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>Peskin et al., 2014</td>
<td>96 60-95</td>
<td>Canada</td>
<td>FB (2nd order), SS, KA</td>
<td>ICC, WM</td>
<td>WM ↔ FB (2nd order)</td>
</tr>
<tr>
<td></td>
<td>Rosenthal &amp; Pexman, 2014</td>
<td>80 48-84</td>
<td>Canada</td>
<td>DD, UL (1st-2nd order)</td>
<td>DCSS</td>
<td>DCSS + other variables ↔ DD</td>
</tr>
<tr>
<td></td>
<td>Rosenqvist et al., 2014</td>
<td>570 36-72</td>
<td>Finland</td>
<td>NEPSY II (EU)</td>
<td>ICC, WM</td>
<td>EF + other variables ↔ EU</td>
</tr>
<tr>
<td></td>
<td>Fizke et al., 2014</td>
<td>62 42-59</td>
<td>Germany</td>
<td>UL (reality known/unknown)</td>
<td>ICC, WM</td>
<td>WM + other variables ↔ UL (reality unknown)</td>
</tr>
<tr>
<td></td>
<td>Morra et al., 2011</td>
<td>130 61-134</td>
<td>Italy</td>
<td>EU</td>
<td>WM</td>
<td>WM ↔ EU</td>
</tr>
<tr>
<td></td>
<td>Henning et al., 2011</td>
<td>195 38-72</td>
<td>Germany</td>
<td>ToM Scale</td>
<td>DCSS</td>
<td>DCSS + other variables ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>Rakocy et al., 2010</td>
<td>80 37-61</td>
<td>Germany</td>
<td>DD, UL, UC</td>
<td>ICC</td>
<td>ICC ↔ DD &amp; UL &amp; UC</td>
</tr>
<tr>
<td></td>
<td>Rakocy et al., 2010</td>
<td>54 61-95</td>
<td>Germany</td>
<td>FB, DD</td>
<td>ICC, WM</td>
<td>ICC ↔ FB &amp; DD</td>
</tr>
<tr>
<td></td>
<td>Guajardo et al., 2009</td>
<td>92 36-71</td>
<td>USA</td>
<td>UL, UC, DC</td>
<td>SH, WM</td>
<td>SH ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>Karpinski &amp; Scullin, 2009</td>
<td>80 36-60</td>
<td>USA</td>
<td>ToM Scale</td>
<td>PL, ICC, WM</td>
<td>EF ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>McAlistier &amp; Peterson, 2006</td>
<td>124 39-69</td>
<td>Australia</td>
<td>UL, AR, PR</td>
<td>ICC, SH, PL</td>
<td>EF ↔ ToM+ other variables in two models</td>
</tr>
<tr>
<td></td>
<td>Melinder et al., 2006</td>
<td>115 36-72</td>
<td>Sweden</td>
<td>ToM Scale</td>
<td>ICC</td>
<td>ICC ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>Perner et al., 2002</td>
<td>56 37-74</td>
<td>Austria</td>
<td>UL (prediction)</td>
<td>ICC</td>
<td>ICC ↔ UL</td>
</tr>
<tr>
<td></td>
<td>Perner et al., 2002</td>
<td>73 33-68</td>
<td>Austria</td>
<td>UL (prediction &amp; explanation)</td>
<td>ICC</td>
<td>ICC ↔ UL (prediction &amp; justification)</td>
</tr>
<tr>
<td></td>
<td>Carlson &amp; Mose, 2001</td>
<td>107 32-39</td>
<td>USA</td>
<td>UL, UC, DC, AR, FB control</td>
<td>ICC, ICH</td>
<td>ICC + other variables ↔ ToM (except deception)</td>
</tr>
<tr>
<td></td>
<td>Keenan et al., 1988</td>
<td>60 48-60</td>
<td>New Zealand</td>
<td>UL</td>
<td>WM</td>
<td>WM + other variable ↔ UL</td>
</tr>
<tr>
<td></td>
<td>Cardillo et al., 2021</td>
<td>70 96-216</td>
<td>Italy</td>
<td>EU, NEPSY-II</td>
<td>WM</td>
<td>N.S.</td>
</tr>
<tr>
<td></td>
<td>Holl et al., 2021</td>
<td>1301 108-136</td>
<td>Germany</td>
<td>EU, IU</td>
<td>ICC, SH, WM</td>
<td>EF ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>Koukli et al., 2017</td>
<td>69 14-44</td>
<td>Greece</td>
<td>UL, MSA/EU</td>
<td>ICC, PL, WM, ICH</td>
<td>Cool EFs + other variables ↔ FB</td>
</tr>
<tr>
<td></td>
<td>(TD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cantin et al., 2016</td>
<td>87 84-120</td>
<td>USA</td>
<td>SS</td>
<td>ICC, WM, SH</td>
<td>ICC &amp; WM / ICC &amp; SH ↔ ToM</td>
</tr>
<tr>
<td></td>
<td>IM-Bolter et al., 2016</td>
<td>444 84-144</td>
<td>Canada</td>
<td>SS</td>
<td>ICC, WM, SH</td>
<td>ICC &amp; WM + other variables ↔ ToM (kids)</td>
</tr>
<tr>
<td></td>
<td>Wang et al., 2016</td>
<td>334 108-192</td>
<td>UK/Hong Kong</td>
<td>MSA, SS</td>
<td>ICC, SH, WM</td>
<td>ICC &amp; SH + other variables ↔ ToM (adolescents)</td>
</tr>
<tr>
<td></td>
<td>Bock et al., 2015</td>
<td>104 84-144</td>
<td>USA</td>
<td>SS, UL 2nd order</td>
<td>WM, SH, ICC</td>
<td>EF + other variables ↔ ToM (Hong Kong)</td>
</tr>
<tr>
<td></td>
<td>Vetter et al., 2013</td>
<td>139 144-264</td>
<td>Germany</td>
<td>EU</td>
<td>ICC, SH, WM</td>
<td>SH + other variables ↔ SS</td>
</tr>
</tbody>
</table>

1. **TD** = Toddlerhood, **EC** = Early Childhood, **MC** = Middle Childhood, **AD** = Adolescence
2. Numbers (1) and (2) next to the authors refer to different experiments reported in the same paper.
3. Ages typed in bold extend beyond the limits of the age group.
4. **ToM** (h vowel: Deception, PR = Pretence, FB = False Belief, UL = Unexpected Location, MSA = Mental State Attribution, EU = Emotion Understanding, KA = Knowledge Access, AR = Appearance-Reality, TK = Think-Know (differential), UC = Unexpected Content, IU = Intentions understanding, PT = Perspective taking, SS = Social Stories, DD = Different Desires).
5. EFs: WM = Working Memory, ICC = Inhibitory Control Coool, SH = shifting/switching/mental flexibility, ICH = Inhibitory Control Hot, DT = Dual Task, PL = planning, CST = card sorting test, PCA = physical causality
6. Symbols referring to the relationship between EFs and ToM: ↔ = variable X is related to variable Y (correlation), → = variable X predicts variable Y (regression), ↔ = variable X predicts variable Y, but variable Y also predicts variable X (regression)
Table 4
Training program protocols

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Authors</th>
<th>Sample</th>
<th>Age (in months)</th>
<th>Training program</th>
<th>Duration</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>Arslan et al., 2018</td>
<td>N=106, EG=72, ACG=34</td>
<td>60-72</td>
<td>1) 3 conditions for EG training (3 FB tasks): (a) feedback (correct/wrong) with explanation (FB justified) (b) feedback without explanation (c) no feedback 2) active control a) EG1: trained in DCCS task b) EG2: trained in FB tasks c) ACG: relative clause tests, and number-conservation tasks</td>
<td>30-45min/session, 2 training sessions, 1/week</td>
<td>WM could not predict ToM performance. The task used for the assessment of ToM was criticized as too simple</td>
</tr>
<tr>
<td></td>
<td>Kloo &amp; Perner, 2003</td>
<td>N=44, EG1=14, EG2=15, ACG=15</td>
<td>36-55</td>
<td>3 training programs: a) Free play b) Sociodramatic Play c) Sociodramatic Play &amp; ToM</td>
<td>15 min/session, 2 training sessions, 1/week</td>
<td>Both DCCS and FB groups significantly improved in both EF and ToM tasks. CG no results.</td>
</tr>
<tr>
<td></td>
<td>Qu et al., 2015</td>
<td>71 divided into 3 conditions</td>
<td>50-74</td>
<td></td>
<td>45min/session, 4 training sessions, 1/week</td>
<td>Children's pretest EF positively predicted the training effect of Sociodramatic Play on children's ToM. Post-EF performance was only related to pre-EF scores.</td>
</tr>
<tr>
<td>MC</td>
<td>Gao et al., 2020</td>
<td>N=96, EG=49, ACG=47</td>
<td>108-120</td>
<td>1) 4 ToM tasks (Misunderstanding, persuasion, double deception, white lies) for EG training 2) similar stories but without including mental state attribution for ACG</td>
<td>40-60min/session, 4 training sessions, 1/week</td>
<td>ToM performance was improved. EF performance did not change. Pre-test EF skills predicted ToM improvement.</td>
</tr>
<tr>
<td></td>
<td>Lecce &amp; Bianco, 2018</td>
<td>N=86, 46 EG, 40 ACG</td>
<td>108-132</td>
<td>1) EG: trained in attributing mental state 2) ACG: physical condition stories (no attribution of mental states)</td>
<td>50 min/session, 4 training sessions, 1/week</td>
<td>ToM skills were improved. WM skills were not improved. Pre-test WM predicted ToM improvement.</td>
</tr>
<tr>
<td></td>
<td>Lecce et al., 2014</td>
<td>N=46 (EF) , EG=25, ACG=21, M=115</td>
<td>45-50 min/session, 4 training sessions, 1/week</td>
<td>1) EG: trained in attributing mental state (use of appropriate verbs and synonyms) 2) ACG: physical condition stories (no attribution of mental states)</td>
<td>Performance in both ToM and EFs tasks during post-training assessment was improved in the EG. The gains in ToM were not related to EFs gains.</td>
<td></td>
</tr>
</tbody>
</table>

1 TD = Toddlerhood, EC= Early Childhood, MC = Middle Childhood
2 EG=experimental group, ACG=active control group
### Longitudinal experimental protocols

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Title</th>
<th>Sample</th>
<th>Age range (months)</th>
<th>Country</th>
<th>Experimental Design</th>
<th>ToM</th>
<th>EF</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD</td>
<td>Kloo et al., 2020</td>
<td>54</td>
<td>18-70</td>
<td>Germany</td>
<td>7 time points of evaluation, between 18-70 months y.o.</td>
<td>UC, UL (1st and 2nd order) (implicit/explicit)</td>
<td>ICC, ICH</td>
<td>Implicit FB and late EF ↔ later explicit ToM</td>
</tr>
<tr>
<td></td>
<td>Müller et al., 2012</td>
<td>82</td>
<td>20-31</td>
<td>Canada</td>
<td>1 assessment/year, 3 times (1h/ses)</td>
<td>DD, EU, PR, UL, UC, PT, DB</td>
<td>ICC, WM, PL</td>
<td>Time 2 ICC + other variables ↔ time 3 ToM (not the opposite)</td>
</tr>
<tr>
<td></td>
<td>Shahaeina et al., 2023</td>
<td>142</td>
<td>24-78</td>
<td>Israel</td>
<td>1 assessment/8 months, 3 times (45min/ses)</td>
<td>UL, UC, EU</td>
<td>WM, ICC</td>
<td>Time 1 EF + other variables ↔ time 2 ToM</td>
</tr>
<tr>
<td></td>
<td>Huang et al., 2022</td>
<td>90</td>
<td>37-64</td>
<td>USA</td>
<td>1 assessment/4 months, 2 time points (15 min/ses)</td>
<td>ToM Scale</td>
<td>WM, ICC, ICH</td>
<td>Time 2 EF &amp; Time 2 ToM + other variables ↔ Time 3 EF</td>
</tr>
<tr>
<td></td>
<td>Brock et al., 2018</td>
<td>354 (T1) 298 (T2)</td>
<td>53-74</td>
<td>USA</td>
<td>1 assessment/year, 2 times</td>
<td>NEPSY II, EU</td>
<td>WM, ICC</td>
<td>Time 2 EF &amp; Time 2 ToM + other variables ↔ Time 3 ToM</td>
</tr>
<tr>
<td></td>
<td>Doenyas et al., 2018</td>
<td>150</td>
<td>36-60</td>
<td>Turkey</td>
<td>1 assessment/year, 1 time (1h/ses)</td>
<td>ToM Scale</td>
<td>ICC</td>
<td>Time 1 EF + other variables ↔ Time 2 ToM</td>
</tr>
<tr>
<td></td>
<td>Devine et al., 2016</td>
<td>137</td>
<td>60-84</td>
<td>UK</td>
<td>1 assessment/4 years, 2 times</td>
<td>UL, UC, SS, EU, MSA</td>
<td>SH, ICC, WM</td>
<td>Time 1 EF + other variables ↔ Time 1 ToM</td>
</tr>
<tr>
<td></td>
<td>Díaz &amp; Farrar, 2018a</td>
<td>78</td>
<td>35-66.8</td>
<td>USA</td>
<td>1 assessment/year, 2 times (40 min/ses)</td>
<td>UL, UC, AR</td>
<td>ICC, SH, STM</td>
<td>Time 2 EF + other variables ↔ Time 2 ToM</td>
</tr>
<tr>
<td></td>
<td>Marcovitch et al., 2015</td>
<td>226</td>
<td>36-60</td>
<td>USA</td>
<td>1 assessment/year, 3 times (2h/ses)</td>
<td>UL, UC, PT, FB (2nd ord)</td>
<td>ICC, SH</td>
<td>Metalinguistic awareness but not EF ↔ Time 2 FB</td>
</tr>
<tr>
<td></td>
<td>McAlister &amp; Peterson, 2013</td>
<td>124</td>
<td>39-80</td>
<td>Australia</td>
<td>2 assessments, 9-15 months interval, 2ses/time,</td>
<td>UC, AR, PR, HE, EU, UL</td>
<td>PL, ICC, WM</td>
<td>Earlier EF + other variables ↔ later ToM performance (not the opposite)</td>
</tr>
<tr>
<td></td>
<td>Jahromi &amp; Stifter, 2008</td>
<td>178</td>
<td>53-65</td>
<td>Canada</td>
<td>1 assessment/year, 1 session/time</td>
<td>UL, FB (2nd order), DD</td>
<td>ICC</td>
<td>Time 1 EF + other variables ↔ Time 2 EF (not the opposite)</td>
</tr>
<tr>
<td></td>
<td>Razza &amp; Blair, 2008</td>
<td>72</td>
<td>42-72</td>
<td>USA</td>
<td>1 assessment/year, 2 times, (45min/ses)</td>
<td>UC, UL, AR</td>
<td>ICC, SH</td>
<td>Time 1 EF + other variables ↔ Time 2 ToM</td>
</tr>
<tr>
<td></td>
<td>Hughes &amp; Ensor, 2007</td>
<td>122</td>
<td>24-48</td>
<td>USA</td>
<td>1 assessment/year, 3 times (2h/ses)</td>
<td>UL, UC, DC, PR, FBlv</td>
<td>ICC, WM, SH</td>
<td>EF time 1 ↔ FBU time 2 but only as a mediator between Time 1 FBU and Time 2 FBU causal relationship</td>
</tr>
<tr>
<td></td>
<td>Schneider et al., 2005</td>
<td>176</td>
<td>36-46</td>
<td>Germany</td>
<td>1 assessment/6 months, 3 times, 3 ses/time</td>
<td>UL, UC, AR</td>
<td>ICC, WM</td>
<td>Time 1 EF + other variables ↔ Time 2 ToM</td>
</tr>
<tr>
<td></td>
<td>Carlson et al., 2004</td>
<td>81</td>
<td>24-39</td>
<td>USA</td>
<td>1 assessment/15 months, 2 times (45min/time)</td>
<td>IU, PT, DD, PR, AR, UC</td>
<td>ICH, ICC</td>
<td>Time 1 EF + other variables ↔ Time 3 EF</td>
</tr>
<tr>
<td></td>
<td>Hughes, 1998</td>
<td>50</td>
<td>39-55</td>
<td>USA</td>
<td>1 assessment/year, 2 times</td>
<td>Time 1: UL, UC</td>
<td>ICC, WM, PL</td>
<td>Time 2 EF + other variables ↔ Time 3 ToM</td>
</tr>
<tr>
<td></td>
<td>Lecce et al., 2017</td>
<td>113</td>
<td>114-128</td>
<td>Italy</td>
<td>1 assessment/year, 3 times</td>
<td>SS, DD, DB, KA (film)</td>
<td>ICC, WM</td>
<td>Time 1 WM + other variables ↔ Time 2 WM</td>
</tr>
<tr>
<td></td>
<td>Austin et al., 2014</td>
<td>3276</td>
<td>72-132</td>
<td>Germany</td>
<td>1 assessment/year, 2 times (50min/ses)</td>
<td>FB 2nd order, SS, MSA, ToM scale</td>
<td>ICC, SH, WM</td>
<td>WM ↔ ToM</td>
</tr>
</tbody>
</table>

1. TD = Toddlerhood, EC= Early Childhood, MC = Middle Childhood, AD = Adolescence
2. Ages typed in bold extend beyond the limits of the age group.
3. ToM (lv=low verbal): DC=Deception, PR=Prentence, FB= False Belief, UL=Unexpected Location, MSA=Mental State Attribution EU=Emotion Understanding, KA=Knowledge Access, AR=Appearance-Reality, TK=Think-Know (differentiation), UC=Unexpected Content, IU=Intentions understanding, PT= Perspective taking, SS=Social Stories, DD=Different Desires, HE=Hidden Emotion
4. EFs: WM=Working Memory, ICC=Inhibitory Control Cool, SH=shifting/switching/mental flexibility, ICH=Inhibitory Control Hot, DT=Dual Task, PL=planning, CST=card shorting test, PCA=physical causality, STM=Short Term Memory
5. Symbols referring to the relationship between EFs and ToM: ↔ = variable X is related to variable Y (correlation), ↔ = variable X predicts variable Y (regression), ↔ = variable X predicts variable Y, but variable Y also predicts variable X (regression).
Brock et al., 2018; Carlson et al., 2004; Doenyas et al., 2018; Huang et al., 2022; Hughes & Ensor, 2007; Hughes et al. 1998; Lecce et al., 2017; Marcovitch et al., 2015; Müller et al., 2012; Shahaeina et al. 2023). Three of these studies also explored the inverse effect of a ToM measure on later EFs but failed to confirm it (Hughes et al., 1998; Marcovitch et al., 2015; Müller et al., 2012). Another three of these studies succeeded in finding a bidirectional effect of some EF measure on later ToM and of some aspect of ToM on later EFs (Austin et al., 2014; Huang et al., 2022; Hughes & Ensor, 2007). One more study (McAlister & Peterson, 2013) only found a ToM-EF effect, while two studies only revealed a mediating role of EFs in the effect between two successive measurements of ToM (Kloo et al. 2020; Razza & Blair, 2008). Finally, two studies failed to confirm any effect of EFs on some aspect of ToM at a later time (Diaz & Farrar, 2018; McAlister & Peterson, 2013) (for more details, see Table 5).

**Cognitive demands of ToM tasks and EFs**

One of the most common criticisms against EFs and ToM developmental interconnection concerns whether this relationship can solely be attributed to the demands of ToM tasks or whether there is a real developmental or causal link between the two. As already noted, expression theory (Moses, 2003) suggests that the association between performance in EFs and ToM may be affected by the nature of ToM tasks.

Indeed, changes in the executive demands of an unexpected location task (e.g., eliminating the need for reality suppression) by decreasing inhibitory load revealed an increase in children’s ToM performance (Fizke et al., 2014; Wang et al., 2012). However, manipulating the verbal demands of ToM tasks (using photographs, signs, or drawings) does not always lead to an enhanced ToM performance. Müller et al. (2005) found a significant correlation between performance in DCCS and an FB task low in verbal demands. However, when a partial correlation controlled for the participants' verbal ability and age, this correlation was no longer significant. Similarly, among the three EFs (IC, WM, SH) assessed by Burnel et al. (2020), only WM and SH were significantly correlated to a ToM task low in verbal demands after controlling for age. Ford et al. (2012) administrated five IC tasks and found a marginal correlation between only one of the IC tasks and a ToM task low in verbal demands. Finally, Sabbagh, Moses et al. (2006) explored the relationship between IC and WM with two low verbal FB tasks, one with photographs and the other with signs. Their results revealed a significant correlation of EFs only with the sign version of the ToM task.

**Discussion**

Only 9 of the 111 studies included in the present systematic review failed to find any kind of relationship between EF and ToM (Arslan et al., 2018; Cardillo et al., 2021; Diaz & Farrar, 2018a; Diaz & Farrar, 2018b; Lecce et al., 2014; Meinhardt-Injac et al., 2020; Scullin & Bonner, 2006; Sudo & Matsui, 2021 -only for the monolingual sample; Talwar et al., 2017). Every aspect of Cognitive and Affective ToM has been linked with at least one of the principal EFs (inhibition, updating, and shifting). The importance of IC in the attribution of FB has been reported and explained in many studies. According to existing data, to perform correctly in ToM prediction and justification tasks, one needs to inhibit and disengage from a salient real-world situation and focus on an abstract mentalistic representation. Therefore, IC is taken to be a prerequisite for FB attribution (Wellman, 2001). Inhibition was also shown to be a crucial factor for the attribution of desires, emotions, knowledge, perspective-taking and deception (e.g., Andrews et al., 2003; Chasiotis et al., 2006; Chu & Minai, 2018; Fizke et al., 2014). However, the relationship between IC and ToM seems to depend on the nature of the task used for measuring EFs. For instance, studies measuring IC through more than one IC task yielded divergent results for each task’s relationship with ToM (e.g., Huang et al., 2022; Ford et al., 2012). As Rey-Mermet et al. (2018) explain, different inhibitory control tasks “do not measure a common underlying construct, but the highly task-specific ability to resolve the interference arising in that task” (p.32) and, thus, relate to FB understanding to different extents (Devine & Hughes, 2014).

The effect of WM on ToM tasks performance has also been addressed by several studies. Simultaneously dealing with two different and controversial realities seems to rely heavily on WM capacity. The vast majority of studies found either a correlation or an effect of WM on one or more ToM tasks. WM has been found to be an important factor for the attribution of beliefs, emotions, and for deception, as well as for the mentalistic understanding of non-literal communication (e.g., Austin et al., 2014; Buac & Kaushanskaya, 2020; Hoyo et al., 2019; Longobardi et al., 2021).
Finally, SH has been found to be linked with the attribution of beliefs, emotions, desires, perspectives, and knowledge (e.g., Bock et al., 2015; Gabriel et al., 2019; Grosse Wiesmann et al., 2017; Miller et al., 2018). The need to shift between different representations seems to be necessary for the correct attribution of a mental state (Bock et al., 2015). SH is related to ToM performance in various developmental phases. However, Gabriel et al. (2019) suggest that it is not until children have reached a certain level of ToM functioning that it becomes the principal executive function for more complicated ToM skills, usually emerging in adolescence. Given the limited number of relevant studies in this age group, this assumption should be further assessed.

Russell et al. (1991) were among the first theorists to propose that self-control is a prerequisite for ToM development. He suggested that endogenous changes in executive control are important for ToM development. Data from longitudinal studies offer insight into the predictive effect of EFs on ToM development. The majority of studies reported positive results on this effect (e.g., Austin et al., 2014; Brock et al., 2018; Huang et al., 2022; Shahgeina et al., 2023), but some studies failed to confirm it (Diaz & Farrar, 2018; McAlister & Peterson, 2013). The training studies also present inconclusive results on the predictive role of EFs on ToM performance. Among the handful of studies available, most studies revealed a positive effect of EFs on ToM (e.g., Gao et al., 2020; Kloo & Perner, 2003; Lecce & Bianco, 2018); however, here again, there have been studies with negative results (Arslan et al., 2018; Lecce et al., 2014).

These contradictory findings could be attributed to the significant heterogeneity among the studies concerning the participants' ages, age-range explored, testing intervals, aspects of EFs and ToM assessed, as well as other contextual factors taken simultaneously into account (e.g., demographic, sociocultural or cognitive factors). Additionally, on several occasions during this review, the concrete tasks used were reported to affect the resulting association between EFs and ToM. This was particularly evident in studies administering simultaneously to the same participants more than one IC task, or more than one EF task, or more than one FB task; results of these tasks, measuring in theory the same construct, were found to differ or even be contradictory (e.g., Burnel et al., 2020; Ford et al., 2012; Kloo & Perner, 2003). Therefore, we assume that another very important factor accounting, at least in part, for the inconclusive findings reported earlier can be the nature and demands of the wide variety of tasks (some of them standardized but many of them research-specific) used to assess either EFs or ToM.

As for the inverse effect of the contribution of ToM in the refinement of the EFs, it appears not to be supported by sufficient data. Despite theories suggesting that an improved understanding of one's mind could potentially lead to greater self-control (Perner, 1991), the number of studies exploring the bidirectional developmental relationship between EFs and ToM is very limited. Results have shown that ToM performance at a particular time point could predict subsequent EFs (Austin et al., 2014; Huang et al., 2022; Hughes & Ensor, 2007; McAlister & Peterson, 2013), although negative results are also present in the literature (e.g., Marcovitch et al., 2015; Müller et al., 2012; Hughes, 1998). Furthermore, among the training protocols focusing on mental state attribution practice, only one resulted in post-training improvement in executive functioning (Kloo & Perner, 2003). However, given the restricted number of heterogeneous studies included in the review, no safe conclusions can be drawn.

Overall, the majority of longitudinal and training studies appear to support the “emergence” account (Moses, 2003), suggesting that EFs contribute to the development of ToM concepts. At the same time, possible effects of decreasing the cognitive demands in FB tasks (e.g., less need for reality suppression, non-verbal tasks) suggest that the “expression” theory remains a noteworthy approach. While reviewing existing literature, adopting a synthetic point of view seems important. An effect of the cognitive demands of ToM tasks does not necessarily translate to a lack of a more direct link between these cognitive functions. This point of view is supported by some studies that found a significant relationship between EFs and ToM even when using ToM tasks low in cognitive or linguistic demands (e.g., Burnel et al., 2020; Müller et al., 2005; Sabbagh, Moses et al., 2006). One could assume that the nature and the demands of a task may mediate an existing relationship of EFs with ToM or even a direct effect of EFs on ToM development. However, the lack of longitudinal studies designed to address this particular hypothesis prevents us from reaching more specific conclusions.
Conclusion

During the last decades, an important number of theorists have tried to explain the link between EFs and ToM (Perner et al., 2002). Researchers have approached this topic by focusing on different parameters, like brain region connections, the cognitive profile of children with neurodevelopmental disorders or of typically developing children (Carruthers & Smith, 1996; Ozonoff et al., 1991) from multiple perspectives and in a variety of ways. Undoubtedly, not only the complex nature of ToM and EFs as theoretical constructs but also the varying nature and cognitive demands of the tasks that are currently used for their assessment obscure the picture of this relationship. In addition, the differing age ranges covered in each study is another important factor that may explain many contradictory results and hinder the generation of a unified account of what one expects to be a complex and dynamic developmental relationship.

Finally, several contextual factors have been reported to mediate the ToM-EFs relationship; cultural context (e.g., Chasiotis et al., 2006; Wang et al., 2016), socio-economic status (e.g., Dicataldo & Roch, 2020; Kara & Selcuk, 2021), bilingualism (e.g., Diaz & Farrar, 2018; Sudo & Matsui, 2021), or the "siblings’ effect" (McAlister & Peterson, 2006) should be further assessed and carefully controlled for in experimental designs.

Overall, this review highlights the need for researchers to adopt clear conceptual definitions and rigorous methodological tools, providing details about the concrete aspects of ToM and EFs they aim to address and making careful and informed choices of the tools they will use accordingly. Contradictory as it may seem, acquiring many very clear partial pieces of the ToM-EFs relationship puzzle could enable us to see the entire picture one day. Importantly, given the asynchronous development of each and every one of these pieces, the implementation of many long-term longitudinal studies, assessing as many aspects of ToM and EFs as possible while controlling for the cognitive demands of the tasks, is deemed necessary so that we can reach a theoretical explanation (vs description) of the dynamic interactions among these complex theoretical constructs.

Funding

The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “1st Call for H.F.R.I. Research Projects to support Faculty Members & Researchers and the Procurement of High-and the procurement of high-cost research equipment grant” (Project Number: HFRI-FM17-13).

References


Cassetta, B. D., Pexman, P. M., & Goghari, V. M. (2018). Cognitive and affective theory of mind and relations with executive functioning in middle childhood. *Merrill-Palmer Quarterly, 64*(4), 514-538. [https://doi.org/10.13110/merrpalmon10982.64.4.0514](https://doi.org/10.13110/merrpalmon10982.64.4.0514)


Miller, S. E., Reavis, R. E., & Avila, B. N. (2018). Associations between theory of mind, executive function, and friendship quality in middle childhood. *Merrill-Palmer Quarterly, 64*(3), 397-426. [https://doi.org/10.13110/merrpalquir1982.64.3.0397](https://doi.org/10.13110/merrpalquir1982.64.3.0397)


Η σχέση μεταξύ Θεωρίας του Νου και Εκτελεστικών Λειτουργιών σε τυπικά αναπτυσσόμενα παιδιά: Μια συστηματική ανασκόπηση

Δήμητρα ΟΙΚΟΝΟΜΑΚΟΥ1, Αλεξάνδρα ΚΑΡΟΥΣΟΥ2, Νικόλαος ΜΑΚΡΗΣ1

1 Τμήμα Δημοτικής Εκπαίδευσης, Δημοκρίτειο Πανεπιστήμιο Θράκης, Αλεξανδρούπολη, Ελλάδα
2 Τμήμα Επιστημών της Εκπαίδευσης στην Προσχολική Ηλικία, Δημοκρίτειο Πανεπιστήμιο Θράκης, Αλεξανδρούπολη, Ελλάδα

ΛΕΞΕΙΣ-ΚΛΕΙΔΙΑ

Θεωρία του Νου (ΘτΝ), Κατανόηση νοητικών αναπαραστάσεων, Εκτελεστικές λειτουργίες, Γνωστική ανάπτυξη, Παιδική ηλικία, Τυπική ανάπτυξη

ΠΕΡΙΛΗΨΗ

Η ανάπτυξη της Θεωρίας του Νου (ΘτΝ) αποτελεί κεντρικό θέμα στη μελέτη της γνωστικής ανάπτυξης. Τις τελευταίες δεκαετίες, ερευνητές που εκπροσωπούν διαφορετικές θεωρητικές προσέγγισες προσπάθησαν να διερευνήσουν την ανάδυση και την ανάπτυξη της κατανόησης νοητικών αναπαραστάσεων και τη δυνητική συμβολή άλλων σύνθετων γνωστικών μηχανισμών σε αυτή την αναπτυξιακή διεργασία. Η σχέση μεταξύ της ΘτΝ και των εκτελεστικών λειτουργιών (ΕΛ) έχει αξιολογηθεί στο πλαίσιο διαφόρων πειραματικών σχεδίων. Παρά τον σημαντικό αριθμό των διαθέσιμων μελετών και την ποικιλία των θεωριών που έχουν προταθεί, δεν υπάρχει ακόμα συναίνεση σχετικά με την επακριβή φύση της σχέσης αυτής. Ο κύριος στόχος της παρούσας ανασκόπησης είναι η συστηματική καταγραφή των σχετικών ερευνητικών αποτελεσμάτων που αφορούν σε παιδιά τυπικής ανάπτυξης. Επιδιώκεται να διαχωριστούν οι παράγοντες που μπορούν ενδεχομένως να εξηγήσουν τα αντιφατικά ευρήματα που αναφέρονται στη βιβλιογραφία. Τα αποτελέσματα της ανασκόπησης μπορούν να γίνουν βάση για την περαιτέρω ανάπτυξη της ΘτΝ και των εκτελεστικών λειτουργιών, ανάλογα με τα αποτελέσματα των μελετών. Η φύση αυτής της σχέσης επιπλέον, οδηγεί και στην επαναλαμβάνοντας συμπίεση της ΘτΝ, αν και στις τρέχουσες μελέτες. Συνεπώς, προτείνεται ότι για την οικοδόμηση μιας ενοποιημένης εικόνας των αναπτυξιακών σχέσεων μεταξύ αυτών των δύο πολυδιάστατων θεωρητικών κατασκευών, είναι απαραίτητη η εκλέπτυνση των εννοιολογικών ορισμών και των μεθοδολογικών προσεγγισεών.