The Impact of Bilingualism on Theory of Mind and Executive Functions in TD and ASD

Eleni Baldimtsi, Eleni Peristeri, Ianthi Maria Tsimpili, and Stephanie Durrleman

1. Introduction

The ability to attribute mental states to oneself and others, also known as Theory of Mind or ToM (Premack & Woodruff, 1978), underlies various social skills (Derksen et al. 2018; Happé, 1993; Mazza et al. 2017; Perner, 1991). Individuals with Autism Spectrum Disorder (ASD) often display ToM difficulties, explaining core impairments in communication and socialization attested in this condition (Baron-Cohen, 1988; Happé, 1994). According to previous research, executive functions (EFs), a set of high-order interacting cognitive processes involved in planning, coordinating and completing goal-directed actions (Miyake et al. 2000), are prerequisites for ToM, in both typically developing (TD) children and those with ASD (Flynn, 2007; Pellicano, 2007). More specifically, EFs such as inhibitory control (Carlson & Moses, 2001; Hughes, 1998), updating (Carlson, Moses & Breton, 2002), cognitive flexibility and working memory (WM) arguably play an important role in the explanation and prediction of a character’s behavior based on their belief (Hughes, Russell, & Robbins, 1993; Russell, 1997). For example, a classical ToM task, i.e. the first order change-of-location false-belief task (see 1), involves EF subcomponents such as WM to track the location of the object, attention shifting to swap from one’s own perspective to that of the protagonist, as well as inhibition to suppress touching the real location of the object so as to point instead to where it is believed to be (Baillargeon, Scott, & He, 2010; Moses, 2001; Nolaker et al. 2018).
Various EF components are reportedly affected in ASD, such as planning (Ozonoff, Pennington, & Rogers, 1991; Ozonoff & Jensen, 1999), mental flexibility (Hughes, Russell, & Robbins, 1994), working memory (Peristeri, Andreou, & Tsimpili, 2017; Schuh & Eigsti, 2012; Tyson et al. 2014) and inhibition (Hughes & Russell, 1993). According to the Executive Dysfunction Hypothesis, difficulties with planning and flexibility (Ozonoff & Jensen, 1999) would explain the development of certain non-social autistic symptoms, such as the restrictive and repetitive patterns of behavior. Impairments in EF have also been argued to hinder individuals with ASD from reasoning about mental states (Pellicano, 2007).

For TD children, it has been claimed that the bilingual experience has advantages for higher cognitive functioning such as EFs (Bialystok, 1999; 2010; 2019; Blom et al. 2014) and ToM (Diaz & Farrar, 2017; Greenberg, Bellana, & Bialystok, 2013; Liberman, Woodward, Keysar, & Kinzler, 2017) as well as for
adults (Morales et al. 2014, though see and Paap & Greenberg, 2013 for criticism and see Lehtonen et al. 2018 for a meta-analysis on this relationship). That the beneficial impact of bilingualism on EF may carry over to ASD has also been suggested by a preliminary study (Gonzalez-Barrero & Nadig, 2017), however more substantial cohorts would be necessary to confirm this to be the case. It also remains to be determined whether ToM boosts can be attested in bilingual ASD. The current study aims to shed light on these issues through an investigation of the effect of bilingualism on ToM and EF in ASD and TD. More specifically, we examine the hypothesis that bilingualism may enhance ToM (i.e. false-belief attribution) and EF (i.e. inhibition, attention-shifting and working memory) not only in TD children, but also in children with difficulties in these realms, namely those with ASD.

2. Method
2.1. Participants

A total of 136 children divided into 4 groups participated in this study: 34 monolingual Greek-speaking children with ASD (ASD-Mono; mean age: 11;7 yrs. (SD: 2.0), age range: 6;9–15;6); 34 bilingual children with ASD (ASD-Bi; mean age:11;6 yrs. (SD: 1.4), age range: 9;1 – 14;5); 35 TD monolingual Greek-speaking children (TD-Mono; mean age: 10;7 yrs. (SD: 1.6), age range: 7;1 – 12;7); and 34 TD bilingual children (TD-Bi; mean age: 11;1 yrs. (SD: 1.7), age range: 8;0 – 13;7). The children were matched across groups for chronological age and gender. Thus, there were no significant differences between groups in age ($F (3, 1336) =1.811, p=.150$). However, there was a significant Group effect in performance IQ (PIQ) as measured by the Greek version of the Weschler Intelligence Scale for Children, (Georgas, Paraskevopoulos, Bezevegis & Giannitsas, 2003) $F (3, 136) = 27.95, p < .001$ (TD-Mono; mean PIQ: 111.11 (SD: 12.92), TD-Bi; mean PIQ: 108.94 (SD: 7.57), ASD-Mono; mean PIQ: 88.7 (SD: 13.82), ASD-Bi; mean PIQ: 98.0 (SD: 16.07). Both groups with ASD had significantly lower PIQ scores than their TD peers ($p < .001$ for the difference between ASDmono and TDmono; $p = .003$ for the difference between ASDbi and TDbi children). There were, crucially, no significant differences either between ASDmono and ASDbi children ($p = .128$) or between TDmono and TDbi children ($p = .916$) in PIQ.

TD children were recruited from mainstream school settings and spoke Greek fluently. Eligibility criteria included normal hearing, no emotional, mental, neurodevelopmental or language impairment as per parent and teacher report. The majority of bilingual TD children were Albanian-Greek speakers. Twenty-eight of them had Albanian as their first language, four had Russian and two Bulgarian.

Children with ASD, both bilingual and monolingual, were recruited for this study with a previous diagnosis of ASD by a licensed clinician (i.e. a child psychiatrist) and their diagnoses were confirmed with the Autism Diagnostic Interview-Revised (Rutter, LeCouteur, & Lord, 2003). In addition, all ASD participants had a non-verbal IQ of at least 70. Furthermore, according to parental
and school reports, participants had no history of language delay. The mother
tongues of the bilinguals with ASD varied, 37 being Albanian-, four Russian- and
two Bulgarian speaking.

2.2. General Procedure

Following informed parental consent, we administered three online tasks in
two sessions to all four groups of children: a low-verbal first-order false-belief
task, an online selective attention (global-local) task, and an online 2-back
updating and working memory task. All tasks were run on a computer using E-
Prime software (Schneider et al., 2012). Tasks were presented in a random order
and children were tested individually at their school or home.

2.3. Experimental Tasks

2.3.1. Low-verbal first-order false-belief task

Stimuli. This ToM test measured comprehension of false-belief (adapted from
Forgeot d’ Arc & Ramus, 2010). Participants watched a series of short videos that
represented different stories having a main protagonist. Each story-based video
included four successive phases. In the beginning phase, common to all
experimental conditions, participants were introduced to the general situation and
the main agent. In the subsequent change phase, five distinct conditions were
presented to the participants. The ‘Mentalistic/Seen change’ and the
‘Mentalistic/Unseen change’ presented a change in the state of the world that was
respectively known or unknown to the main protagonist. The ‘Mentalistic/No
change’ condition, in contrast, implied no change whatsoever. The
‘Mechanistic/Unseen Change’ condition involved an event that implied a change
unrelated to an individual’s mental state. The ‘Mechanistic/No change’
condition, like the ‘Mentalistic/No change’ condition implied no change
whatsoever. In the suspense phase (which was identical to all conditions)
participants saw the main protagonist coming to the front. Finally, the end phase
had two alternative endings, the ‘Mentalistic end’ that required the attribution of
the protagonist’s belief in the given situation and so to track and predict his
behavior, and the ‘Mechanistic end’ in which participants needed to understand
and predict the outcome of an event in the psychical world, without having to infer
a belief or an intention of the main agent. The possible ends in the mentalistic
conditions of the task were two videos depicting opposite behaviors, while the
possible ends in the mechanistic conditions were two videos showing different
resulting states (see Forgeot d’ Arc & Ramus, 2010 for a detailed description
of the task’s design). The task contained ten different stories in five conditions, with
two different endings coming to a sum of one hundred video-based stories.
Participants completed the task during two sessions within a week. Before the
actual experiment children completed a familiarization training of ten trials.

Procedure. After completion of each story, participants saw a question mark
[?] in the middle of the computer screen and were asked whether the end of the
story was logical or not. Children had to decide as quickly and accurately as possible and then to press the appropriate key in the keyboard (a green- or a red-colored [√] button, respectively). Response times (i.e. time in ms from the appearance of the question mark to the child’s button-press) and accuracy (%) of judgments were recorded via E-Prime software (Schneider, Eschman, & Zuccolotto, 2012).

Data analysis. We analyzed accuracy on both the ‘Mentalistic Unseen change’, which required the attribution of a false belief, and the ‘Mechanistic Unseen change’ condition, in which participants did not need to infer a belief. Accuracy scores were computed as follows: first, calculation of the percentage means of correct and false decisions, and then subtraction of the wrong answers from the percentage mean of correct decisions. Reaction times on correct decisions were also analyzed.

2.3.2. Online visual attention-switching global-local task

Materials. This task was modeled after Navon (1977). Stimuli were large (global) geometrical figures like circles, Xs, triangles and squares made of smaller (local) same or different figures. The figures were designed with the AutoCAD 2012 software and were identical in size (the scale from global to local was 1:10).

Procedure. The task included three blocks/conditions (a global to global, a local to local and a local to global or global to local/switch block trial). Each child was presented with either the global or the local stimuli and had to indicate which shape each figure was made of and press the appropriate button on the keyboard. More specifically, children were trained to press ‘1’ for circle, ‘2’ for an X, ‘3’ for a triangle, and ‘4’ for a square depending on the number of lines needed to form the shape. Each condition consisted of 48 trials evenly divided into same (congruent) and different (incongruent) trials. The stimuli presented to the participants were randomized. Moreover, to avoid any response bias effect, half of participants were presented with the ‘Global-first’ condition first and the other half with the ‘Local-first’ condition first. Written and verbal instructions as well as a familiarization phase of 16-items preceded each task condition. Local to global (and vice versa) switch costs in both RT and accuracy (%) separately for congruent and incongruent trials were recorded (for a full description of the task and the scoring see Peristeri, Andreou, Tsimpli & Durrleman, 2019).

2.3.3. Updating (2-back) task

Materials. In this task, children were presented with a sequence of digits (2, 5, 7, 8) on the computer screen. The task contained a total of 20 to-be-responded to (target) digits and 60 trials on the whole. Children had to remember if the digit they saw on the computer screen was the same as the one presented two digits back in the sequence, and if it was the same, to press a pre-specified key (‘J’) with their index finger. They were instructed to press no button for non-target digits. Children started with a familiarization of 20 trials. Each trial consisted of a black,
12mm-long digit that was presented for 500 msecs, followed by a blank page that lasted 2500 msecs, after which the next digit stimulus was presented. An accuracy composite score was computed for each child by calculating the percentage means of correct hits and false hits (i.e., by dividing the number of correct hits by the 20 target trials and the number of false hits by the 40 non-target trials, respectively) and then by subtracting the percentage mean of false hits from the percentage mean of correct hits. Reaction times on correct hits were also recorded.

3. Results

3.1. Low-verbal first-order false-belief task

Table 1 provides descriptive statistics for the groups’ accuracy and RT performance in the ‘Mechanistic Unseen change’ and ‘Mentalistic Unseen change’ condition. Prior to statistical analyses, subject-by-subject RT on judgments which were smaller than 250 msecs and over 2 SDs from the mean RT for each condition (Mechanistic, Mentalistic) were considered outliers and replaced by the mean.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mechanistic Unseen change</th>
<th>Mentalistic Unseen change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy (%) (SD)</td>
<td>Accuracy (%) (SD)</td>
</tr>
<tr>
<td></td>
<td>RT (in msecs) (SD)</td>
<td>RT (in msecs) (SD)</td>
</tr>
<tr>
<td>ASD-Mono</td>
<td>76.5 (17.2)</td>
<td>45.6 (16.9)</td>
</tr>
<tr>
<td>(N = 34)</td>
<td>2729 (544)</td>
<td>2129 (734)</td>
</tr>
<tr>
<td>ASD-Bi</td>
<td>83.0 (17.6)</td>
<td>74.8 (16.3)</td>
</tr>
<tr>
<td>(N = 34)</td>
<td>2403 (605)</td>
<td>2746 (730)</td>
</tr>
<tr>
<td>TD-Mono</td>
<td>77.7 (11.6)</td>
<td>79.4 (8.2)</td>
</tr>
<tr>
<td>(N = 35)</td>
<td>2639 (487)</td>
<td>2301 (391)</td>
</tr>
<tr>
<td>TD-Bi</td>
<td>80.6 (15.7)</td>
<td>75.8 (9.5)</td>
</tr>
<tr>
<td>(N = 34)</td>
<td>2527 (716)</td>
<td>2382 (474)</td>
</tr>
</tbody>
</table>


A mixed-design ANCOVA was conducted with Group (ASD-Mono, ASD-Bi, TD-Mono, TD-Bi) as the between-subjects factor, Trial-type (mechanistic, mentalistic) as the within-subjects factor, and age and PIQ as covariates. We first report the results of the RT measure. The repeated measures analysis revealed neither a significant group effect, $F(3, 131) = .159, p = .924, \eta^2 = .01$, nor a significant age, $F(1, 131) = .222, p = .638, \eta^2 = .00$, and PIQ effect, $F(1, 132) = .402, p = .527, \eta^2 = .00$. The two-way interaction between trial-type
(mechanistic, mentalistic) and Group was not significant either, $F(3, 131) = 1.435, p = .236, \eta^2 = .03$. Regarding accuracy, the analysis revealed a significant Group effect, $F(3, 131) = 14.099, p < .001, \eta^2 = .24$. Subsequent post-hoc (Bonferroni) tests showed that the ASD-Mono group’s overall accuracy score was significantly lower than the rest of the experimental groups ($p < .001$ for all comparisons). The age, $F(1, 131) = 1.088, p = .298, \eta^2 = .01$, and the PIQ effect, $F(1, 131) = .132, p = .717, \eta^2 = .00$, were not significant. The two-way interaction between trial-type (mechanistic, mentalistic) and Group was statistically significant, $F(3, 131) = 10.261, p < .001, \eta^2 = .19$. Subsequent paired samples $t$-tests revealed that mentalistic trials were responded significantly more erroneously than mechanistic trials for the ASD-Mono group only ($t(33) = 6.433, p < .001$. This difference was not replicated for any of the other experimental groups ($t(33) = 2.369, p = .093$ for ASD-Bi, $t(34) = .488, p = .629$ for TD-Mono, and $t(33) = 1.436, p = .160$ for TD-Bi children).

3.2. Selective attention global-local task

Table 2 provides descriptive statistics for the groups’ accuracy and RT performance in the ‘global-to-local’ and ‘local-to-global’ switching condition. Prior to statistical analyses, subject-by-subject RT on responses which were smaller than 250 msecs and over 2 $SD$s from the mean RT for each condition were considered outliers and replaced by the mean.

<table>
<thead>
<tr>
<th>Group</th>
<th>Global-to-local</th>
<th>Local-to-global</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy (%)</td>
<td>RT (in msecs)</td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(SD)</td>
</tr>
<tr>
<td>ASD-Mono</td>
<td>11.8 (15.8)</td>
<td>370 (290)</td>
</tr>
<tr>
<td>($N = 34$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD-Bi</td>
<td>23.6 (10.5)</td>
<td>443 (382)</td>
</tr>
<tr>
<td>($N = 34$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD-Mono</td>
<td>17.4 (6.5)</td>
<td>1397 (832)</td>
</tr>
<tr>
<td>($N = 35$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD-Bi</td>
<td>28.2 (15.9)</td>
<td>1217 (628)</td>
</tr>
<tr>
<td>($N = 34$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


A mixed-design ANCOVA was conducted with Group (ASD-Mono, ASD-Bi, TD-Mono, TD-Bi) as the between-subjects factor, Switch level (local-to-global, global-to-local) as the within-subjects factors, and age and PIQ as the
covariates. The analysis on the groups’ RTs revealed a significant Group effect, $F(3, 131) = 4.902, p < .001, \eta^2 = .10$. Subsequent post-hoc (Bonferroni) tests showed that both TD groups’ overall switching costs were significantly higher compared to the ASD groups ($p = .013$ and $p = .013$ for the difference between TD-Mono and ASD-Mono & ASD-Bi groups, respectively, and $p = .011$ and $p = .003$ for the difference between TD-Bi and ASD-Mono & ASD-Bi groups, respectively). The age effect, $F(1, 131) = .339, p = .561, \eta^2 = .00$, and the PIQ effect, $F(1, 131) = .213, p = .645, \eta^2 = .00$, were not found to be significant. The two-way interaction between attention level (global-to-local, local-to-global) and Group was statistically significant, $F(3, 131) = 9.727, p < .001, \eta^2 = .18$. Subsequent paired samples $t$-tests revealed that local-to-global interference was stronger than global-to-local interference for the ASD-Mono group ($t(33) = 7.995, p < .001$), while the opposite pattern was observed for ASD-Bi children ($t(34) = 4.065, p < .001$). There was no significant difference between global-to-local and local-to-global costs in RT for either TD-Mono or TD-Bi children ($t(34) = 1.397, p = .172$ and $t(34) = 1.483, p = .147$, respectively).

The analysis on the groups’ accuracy data revealed a significant Group effect, $F(3, 131) = 6.133, p < .001, \eta^2 = .12$. Subsequent post-hoc (Bonferroni) tests showed that the ASD-Mono group’s overall attention switching score was significantly higher than the rest of the experimental groups ($p = .031, p < .001$, and $p = .003$ for the difference with ASD-Bi, TD-Mono and TD-Bi children, respectively). Age, $F(1, 131) = 1.296, p = .257, \eta^2 = .01$, and PIQ, $F(1, 131) = .141, p = .708, \eta^2 = .01$, were not found to be significant. The two-way interaction between attention level (global-to-local, local-to-global) and Group was statistically significant, $F(3, 131) = 20.644, p < .001, \eta^2 = .32$. Subsequent paired samples $t$-tests revealed that local-to-global interference was stronger than global-to-local interference for the ASD-Mono group ($t(33) = 6.123, p < .001$), while the opposite pattern was observed for both TD-Mono and TD-Bi children ($t(34) = 4.634, p < .001$ and $t(33) = 4.908, p < .001$, respectively). There was no significant difference between global-to-local and local-to-global costs in accuracy for the ASD-Bi group ($t(34) = .934, p = .357$).

### 3.3. Updating (2-back) task

Table 3 provides descriptive statistics for the groups’ accuracy and RT performance in the 2-back task. Prior to statistical analyses, subject-by-subject RTs on responses which were smaller than 250 msecs and over 2 $SD$s from the mean RT for each condition were considered outliers and replaced by the mean.
Table 3. Groups’ mean accuracy scores and RTs on correct responses (and SDs) per group in the 2-back task

<table>
<thead>
<tr>
<th>Group</th>
<th>Accuracy (%) (SD)</th>
<th>RT (in msecs) (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD-Mono</td>
<td>25.1 (15.8)</td>
<td>460 (184)</td>
</tr>
<tr>
<td>(N = 34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD-Bi</td>
<td>35.3 (12.9)</td>
<td>448 (126)</td>
</tr>
<tr>
<td>(N = 34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD-Mono</td>
<td>19.4 (10.8)</td>
<td>420 (130)</td>
</tr>
<tr>
<td>(N = 35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD-Bi</td>
<td>28.7 (13.3)</td>
<td>353 (81)</td>
</tr>
<tr>
<td>(N = 34)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


A one-way ANCOVA analysis was conducted with Group (ASD-Mono, ASD-Bi, TD-Mono, TD-Bi) as the between-subjects factor, accuracy and RT as the dependent variables, and age and PIQ as the covariates. The analysis on accuracy revealed a significant Group effect, \( F(3, 136) = 2.743, p = .046, \eta^2 = .06 \), which stemmed from the fact that ASD-Bi children scored significantly higher than their TD-Mono peers \( p = .036 \). The age, \( F(1, 136) = 2.956, p = .088, \eta^2 = .02 \), and the PIQ effect, \( F(1, 136) = .229, p = .633, \eta^2 = .00 \), were not found to be significant. Considering RT, the same analysis revealed a significant Group effect, \( F(3, 136) = 3.902, p = .01, \eta^2 = .08 \), which stemmed from the fact that TD-Bi children were considerably faster than both ASD-Mono \( p = .007 \) and ASD-Bi children \( p = .025 \). The age, \( F(1, 136) = .326, p = .569, \eta^2 = .00 \), and the PIQ effects, \( F(1, 136) = .152, p = 697, \eta^2 = .00 \), were not found to be significant.

4. Discussion

This study focused on bilingualism, ToM and cognitive control (i.e. EF) in children with and without ASD. While many studies have examined the impact of bilingualism on EF and ToM in TD children (Bialystok, 1999; Bialystok & Senman, 2004), research on these domains in individuals with ASD are quasi-inexistent, with only one study on EFs in bilingual ASD involving extremely limited sample sizes (e.g. Gonzalez-Barrero & Nadig, 2017), and absolutely no study exploring the effects of bilingualism on ToM in this condition. We address this lacuna by examining the impact of bilingualism in ASD on a low-verbal, false-belief task (i.e. ToM), as well as on two low-verbal tasks assessing visual switching attention and updating (i.e. EF). Bilinguals with and without ASD were expected to perform better than monolinguals on both the ToM and EF tasks.

The results revealed that bilingual children with ASD outperform their monolingual ASD peers on ToM, indicating that bilingualism may alleviate
challenges caused by autism in the area of mentalizing. Several studies have reported the positive effect of bilingualism on ToM (e.g. Barac et al. 2014; Berguno, & Bowler, 2004; Kovács, 2012, for a meta-analysis see Schroeder, 2018) and the findings of this study showed similar bilingual advantages in ASD children, a population with specific challenges in this realm.

The analysis of children’s performance in the low-verbal visual attention switching EF task also showed significant differences between the two groups with ASD. More specifically, bilingual children with ASD exhibited a less detail-focused processing style compared to their monolingual peers with ASD which was mainly evident in their less strong local-to-global interference (in both how accurate and how fast their attention switched in the local-to-global condition). These results indicate that bilinguals with ASD thus performed more similarly to their TD peers, which was also reported in a preliminary similar study with ASD children by Baldimtsi and colleagues (Baldimtsi, Peristeri, Tsimpli & Nicolopoulou, 2016).

Finally, the two-back updating EF task also revealed a positive bilingualism effect due to the superior performance of the ASD bilinguals as compared to their monolingual TD peers. In addition, the bilingual TD group’s RT performance was found to be significantly better compared to the monolingual TD children and monolingual and bilingual ASD, respectively.

Overall, the findings from the current study show a significant positive effect of bilingualism on both TD and ASD groups. More specifically, both bilingual groups’ better performance on low-verbal ToM and on inhibition, attention-shifting and working memory tasks supports the hypothesis that bilinguals’ ability to constantly inhibit one of their languages during communication may contribute to a generally enhanced inhibitory control system. This may consequently result in more effective inhibition of one’s personal knowledge during a ToM task (Bialystok, 2001; Buac & Kaushanskaya, 2019). Future work should seek to elucidate whether indeed bilinguals’ enhanced EF abilities underlie their boost in ToM performance, both in TD children (Diaz & Farrar, 2017; Nguyen & Astington, 2014) as well as in those with ASD, as has been claimed for their monolingual peers (Ozonoff & Jensen, 1999; Pellicano, 2007; 2010).

References


