Investigating the development of core cognitive skills in autism: a 3-year prospective study

Elizabeth Pellicano

Pellicano, E. (2010). The development of core cognitive skills in autism: a 3-year prospective study. *Child Development, 81,* 1400-1416.

Abstract

This longitudinal study tested the veracity of one candidate multiple-deficits account of autism by assessing 37 children with autism (*M* age=67.9 months) and 31 typical children (*M* age=65.2 months) on tasks tapping components of theory of mind (ToM), executive function (EF), and central coherence (CC) at intake and again 3 years later. As a group, children with autism showed poor false-belief attribution, planning ability and set-shifting, together with enhanced local processing at both time-points. At an individual level, however, the profile was far from universal at either intake or follow-up. Moreover, autistic children demonstrated significant changes over time in ToM and EF, but not CC, over the 3-year period. The challenges these findings pose for a multiple-deficits account are discussed.

The past few decades have seen considerable research efforts devoted to isolating a single underlying cognitive deficit that could provide a unifying explanation for the "triad of impairments" in autism – the severe difficulties in reciprocal social interaction and communication, and restricted interests and repetitive behaviors (American Psychiatric Association [APA], 1994). Three cognitive theories in particular have dominated the field: (1) the theory of mind (ToM) hypothesis, which claimed that autism is caused primarily by a specific inability to impute mental states to oneself and to others (Baron-Cohen, Leslie, & Frith, 1985; see Baron-Cohen, Tager-Flusberg, & Cohen, 2000, for review); (2) the executive dysfunction hypothesis, which proposed that the symptoms of autism are a result of a primary problem in the planning and execution of complex actions (Hughes & Russell, 1993; Ozonoff, Pennington, & Rogers, 1991; see Hill, 2004, for review); and (3) "weak" central coherence (CC) theory, which posited that inherent to autism is an unusual tendency to focus on individual, local elements rather than global wholes (Frith, 1989; Happé & Frith, 2006; see also Mottron & Burack, 2001, for a competing account).

The notion of a single cause at the cognitive level of analysis has, however, been challenged on two main grounds. First, all three influential "single-deficit" models have struggled to satisfy key criteria for an explanatory account of autism, namely that the putative deficit should be "universal, specific, and necessary and sufficient to cause the symptoms of the disorder" (Pennington & Ozonoff, 1996, p. 57), and show persistence or stability over time (Ozonoff & McEvoy, 1994; Rutter, 1983) (see Rajendran & Mitchell, 2007, for review). Second, recent research has called into question the very premise underpinning the search for a core cognitive deficit: that the behaviors central to autism cluster together more often that what would be expected by chance as a result of shared etiology (Happé & Ronald, 2008; Happé, Ronald, & Plomin, 2006b; Mandy & Skuse, 2008). Investigations of the broader phenotype of autism have demonstrated that among those relatives of individuals with autism who display subclinical autistic features, the majority of these family members show behaviors in isolated

symptom domains (e.g., social difficulties in the absence of repetitive behaviors, etc.) rather than a combination of these symptoms, albeit in more subtle form than in autism itself (Bolton, Murphy, MacDonald, Whitlock, Pickles, & Rutter, 1994; Piven et al., 1997). Furthermore, recent findings from a population-based study investigating autism-related traits in over 3000 7- to 9-year-old typically developing twin-pairs found that the cross-trait genetic correlations were surprisingly modest-to-low both across the general population and in children lying at the extreme end of the distributions, suggesting that largely independent genes may be operating on each aspects of the triad of impairments in autism (Ronald et al., 2006). These authors also showed that a considerable number of children displayed behavioral difficulties in isolation (e.g., social difficulties only, communication problems alone). Both sets of findings provocatively suggest that the triad of impairments may in fact be dissociable, raising the possibility of distinct causes for each aspect of the behavioral phenotype.

In light of these challenges, Happé and colleagues have argued that "at the cognitive level, as at the symptom/behavioral and genetic levels, autism may be characterized by fractionable impairments" (Happé & Ronald, 2008, p. 296). Consequently, they have called for researchers to "abandon the attempt to find a single cognitive explanation" (Happé et al., 2006b, p. 1219) in favor of an explanation encompassing coexisting atypicalities in *multiple* cognitive domains. These authors offered one candidate account of their own in which a combination of cognitive atypicalities in three core domains – ToM, executive function (EF), and CC – exists in autism, where each atypicality underpins a distinct aspect of the behavioral phenotype (see also Frith, 2003; Happé & Frith, 2006). Since these atypicalities are held to be largely *independent* of each other both at the phenotypic and genetic levels, autistic symptomatology is therefore viewed as the result of multiple, *primary* cognitive atypicalities.

Recent evidence supports this view. Pellicano, Maybery, Durkin, and Maley (2006) found that young children with an autism spectrum disorder (ASD) (n=40) as a group did in fact demonstrate the cognitive profile proposed by Happé and colleagues, including difficulties

in aspects of ToM (false-belief understanding) and EF (planning, cognitive flexibility) accompanied by weak CC (enhanced local information processing), relative to age- and ability-matched typical children. This support at the group level was somewhat tempered, however, by a failure to demonstrate the presence of this specific cognitive profile in each child with ASD.

One crucial question is how this specific cognitive profile manifests itself over time. Autism is generally understood as a pervasive, lifelong condition (APA, 1994), and therefore any putative explanation of the pathogenesis of autism should be able to account for the continuities and discontinuities that take place across the life span. Despite the wealth of research on specific cognitive skills in autism, however, there have been disappointingly few empirical studies tracking the *development* of such skills in autism – even within a single-deficit framework. The overarching goal of the current study therefore was to investigate longitudinally the multiple-deficits account proposed by Happé et al.'s (2006b) focusing specifically on questions regarding developmental persistence – both at the group level but also at the individual level.

Before describing this study it is worth considering briefly the findings of the five existing longitudinal studies. Remarkably, only four longitudinal investigations have traced developmental changes in ToM. Holroyd and Baron-Cohen (1993) followed 17 of the 20 children with autism described originally in Baron-Cohen et al. (1985) and re-assessed their false-belief understanding 7 years later. No overall gains in participants' ToM skills were found: only three individuals passed the simple 'Sally-Ann' test at follow-up (including two of the four children who had succeeded on this task 7 years earlier), and no individual passed a more advanced ToM test. Ozonoff and McEvoy (1994; Expt 2) also reported no significant changes in the proportion of cognitively-able autistic adolescents passing and failing simple and more advanced ToM tasks over a 3-year period. Indeed, the scores of only 4 individuals with autism improved over time. The authors therefore concluded that, on the whole, ToM abilities progress little with development.

Two recent studies, however, paint a more positive picture. Steele, Joseph, and Tager-Flusberg (2003) found significant improvements in autistic children's ToM skills over a 1-year period, with initial vocabulary level, rather than chronological age, predicting a significant (yet small; 3%) proportion of the variance in later ToM scores. Such gains, however, were driven largely by improvements on tasks tapping early-emerging ToM skills (understanding of pretense and desire) rather than on tasks of false-belief understanding. Serra, Loth, van Geert, Hurkens, and Minderaa (2002) also reported significant gains in a range of ToM skills over a 6-month period in preschool children with a mild variant of autism (pervasive developmental disorder – not otherwise specified; PDD-NOS), even though children with PDD-NOS developed at a slower rate than younger typically developing children. Their analysis, however, examined changes over time for an omnibus ToM score rather than for individual ToM task scores rendering it unclear whether such gains occurred for false-belief understanding per se.

To date, only two studies have investigated the developmental course of EF. In Ozonoff and McEvoy's (1994; Expt 1) 3-year follow-up study, non-autistic adolescents with learning disability showed significant gains in planning efficiency (on the Tower of Hanoi) and made fewer perseverative responses (on the Wisconsin Card Sort Test [WCST]), but concomitant improvements were not evident in adolescents with autism. They concluded that, akin to their findings on ToM, difficulties in EF also persist over time. Griffith, Pennington, Wehner, and Rogers (1999) reported similar findings in preschoolers with autism on a spatial reversal task, a test of cognitive flexibility. Despite there being a trend for children with developmental delay to commit fewer perseverative errors at follow-up, autistic children's performance did not change significantly within the space of one year.

The findings from these longitudinal studies indicate that there are, on the whole, few changes in ToM or EF over time, implying that there might be a ceiling on the extent to which such abilities can develop in people with autism. Yet there is wide variability in the age and ability of participants tested within and across the five studies, which not only makes

comparisons between these few studies difficult, but also can potentially swamp longitudinal effects (i.e., improvements) within studies. Also, with the exception of Ozonoff and McEvoy's (1994) study, there has been no investigation of developmental changes in *multiple* cognitive domains; it is unclear, therefore, whether all core aspects of the cognitive phenotype (cf. Happé et al., 2006b) take a similar developmental course. Indeed, no study has examined longitudinally the development of a local processing bias in children with autism, and therefore the progression of weak CC remains unknown.

The limited knowledge on the developmental course of each of the three cognitive domains presents an initial difficulty for a multiple-deficits account of autism. A complete picture of the development of the condition requires an understanding of the potential changes that take place over time within each domain, in addition to knowledge of the factors that influence such changes. This longitudinal study sought to take a crucial step in this direction. To this end, Pellicano et al.'s (2006) samples of cognitively-able children with ASD and typically developing children were followed prospectively and re-evaluated three years later on similar measures tapping key components of ToM, EF, and CC.

The primary aims of this study were twofold. The first aim was to test whether Happé et al.'s (2006b) specific cognitive profile persisted across time both at the group level and at the individual level. If the putative cognitive profile – difficulties in ToM and EF combined with weak CC – is indeed central to autism, then the group of children with autism should show atypicalities in ToM, EF and CC relative to a group of typically developing children both at intake and at follow-up (cf. Ozonoff & McEvoy, 1994). Furthermore, if the specific combination of atypicalities in ToM, EF, and weak CC underlies autistic symptoms as Happé et al. (2006b) proposed, then this profile should be present in all, or almost all, autistic children at both time points.

The second aim of this study was to determine the nature and extent of developmental changes in ToM, EF, and CC, and to identify potential predictors of cognitive change. The

majority of existing studies has failed to demonstrate developmental improvements in false-belief understanding and aspects of EF. It was expected, however, that longitudinal changes might be more readily detected in a sample of children with autism who were selected initially to be of at least average intellectual functioning and who fell within a restricted age range. Furthermore, language has been implicated in theoretical models of the development of both ToM (e.g., de Villiers, 2001) and EF (e.g., Russell, 1996), and has been linked empirically to autistic children's skills in ToM (e.g., Happé, 1995; Steele et al., 2003), EF (e.g., Liss et al., 2001; Pellicano, 2007), and CC (Morgan, Maybery, & Durkin, 2003; Pellicano et al., 2006). It was hypothesized therefore that, for children with autism, individual differences in general cognitive ability, particularly verbal ability, would be influential in predicting developmental change in all three cognitive domains.

Method

Participants

A total of 90 children participated in the initial study (Time 1): 45 children with an ASD (40 boys) and 45 typically developing children (37 boys). Briefly, all children in the ASD group had received an independent clinical diagnosis of either autism (n = 31), PDD-NOS (n = 12), or Asperger syndrome (n = 3), according to DSM-IV criteria (APA, 1994) and met either full or partial criteria on the Autism Diagnostic Interview – Revised (ADI-R; Lord et al., 1994). These children were recruited via community contacts (see Pellicano et al., 2006, for details). At intake, the ASD group had a mean age of 67.2 months (SD = 10.51, range = 49 – 88) and the typically developing group had a mean age of 65.1 months (SD = 11.80, range = 48 – 88). No child had a coexisting medical (e.g., epilepsy) or developmental (e.g., attention deficit/hyperactivity disorder) condition as reported by parents, or obtained a verbal or nonverbal IQ score below 80, as assessed by the Peabody Picture Vocabulary Test – Third Edition (PPVT-III; Dunn & Dunn, 1997) and the Leiter International Performance Scale – Revised version (Leiter-R; Roid & Miller, 1997), respectively. Note that, unlike more standard IQ tests, which assess both

receptive and expressive language ability, the measure used here to index verbal IQ assessed children's receptive vocabulary skills alone. Children with autism were recruited through early intervention agencies, speech and language therapists, pediatricians, and local support groups in the south-west region of Western Australia. The majority of children were White, and the parents were of mixed socioeconomic backgrounds, although specific data on socioeconomic status and educational attainment levels were not recorded.

Attempts were made to contact all families approximately three years after their involvement in the initial study to invite them to participate in the follow-up study (Time 2). Sixty-eight families were re-traced and gave consent for re-assessment (76% of the original sample) (see Table 1). The mean time between the initial and follow-up studies was 32.9 months for the ASD group and 32.7 months for the typically developing group. Thirty-seven out of 45 children with ASD (82%) were available for further testing. Of the 8 families who were unavailable for reassessment, 2 had relocated to a different state, while 6 families had moved and were untraceable. The 37 children with autism (33 boys) in the follow-up group were not significantly different from those who were not reassessed (n = 8) in terms of chronological age, $F(1, 44) = .28$, $p = .60$, verbal ability, $F(1, 44) = .63$, $p = .43$, nonverbal ability, $F(1, 44) = 1.27$, $p = .27$, or total algorithm score on the ADI-R, $F(1, 44) = 1.90$, $p = .18$. The follow-up group consisted of 26 children with autism, 9 children with PDD-NOS, and 2 children with Asperger syndrome, all of whom met full or partial criteria on the ADI-R conducted at Time 1 (social domain: M = 17.81, SD = 5.74; communication domain: M = 13.22, SD = 4.37; repetitive behaviors domain: M = 6.68, SD = 2.42). The Autism Diagnostic Observation Schedule – Generic (ADOS-G; Lord, Rutter, DiLavore, & Risi, 1999) was administered at follow-up to index these children's current socio-communicative symptoms (Social Interaction score: M = 8.54, SD = 3.51; Communication score: M = 4.11, SD = 2.22; Total algorithm score: M = 12.62, SD = 4.98).

Of the initial sample of 45 typically developing children, who were recruited initially from several local mainstream preschools and primary schools, 31 (69%) took part in the follow-up study (see Table 1). One family had relocated, four families declined to participate, and 10 families were untraceable. The 31 typically developing children (25 boys; 81%) involved in the follow-up study did not differ significantly from the 14 children who did not participate in terms of age, $F(1, 44) = .02$, $p = .87$, verbal ability, $F(1, 44) = 1.66$, $p = .20$, or nonverbal ability, $F(1, 44) = 1.78$, $p = .19$, at intake. Parents of typically developing children completed the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) to ensure that no child in this group displayed clinically-significant levels of autistic symptomatology; all scores were well below the cut-off score of 15 for autism at Time 1 ($M = 4.00$; $SD = 3.32$) and Time 2 ($M = 3.50$; $SD = 2.85$).

Insert Table 1 about here

Measures

Children completed a variety of measures assessing key aspects of cognition at each time point.

Verbal and nonverbal ability. At both time points, the PPVT-III (Dunn & Dunn, 1997) was used to assess children's receptive vocabulary and four subtests of the Leiter-R (Roid & Miller, 1997) were used to index nonverbal ability: *Matching* (a match-to-sample task using pictures of objects and abstract patterns), *Associated Pairs* (an associative memory task in which children were required to establish associations between pairs of pictured objects), *Forward Memory* (a visual short-term memory task that involved children copying the examiner's pointing sequence), and *Attention Sustained* (a visual attention task, which entailed identifying multiple target stimuli amongst distractor stimuli). Standard scores are reported in Table 1, but raw scores are used in correlational and regression analyses since such scores have not been adjusted for age, and therefore are more sensitive to developmental change.

Theory of mind. At each time point, children were presented with three standard false-belief prediction tasks to index ToM ability. For the *first-order unexpected-contents task* (based on Perner, Leekam, & Wimmer, 1987), children were shown 3 different boxes (e.g., milk carton) containing unexpected contents (e.g., rubber bands). After the lid of each box was replaced, they were asked to recall their own false belief (e.g., "Before you looked inside, what did you think was in the box?") and to predict another's false belief (e.g., "If I show this box to Mum, what will she think is inside?). Children also answered a control question (e.g., "What is in the box really?") following each false belief question. For the *first-order unexpected-transfer task* (based on Baron-Cohen et al., 1985), children were shown six different scenarios in which one character either displaced or substituted another character's object. At the end of each trial, children were asked a critical belief question (e.g., "Where will Sarah look for her apple?"), and two control questions (e.g., "What is really in the bag?" and "What was in the bag in the beginning?"). For the *second-order unexpected-transfer task* (based on Perner & Wimmer, 1985), children were shown two displacement scenarios similar to those in the first-order unexpected-transfer task, though this time they witnessed the protagonist watching the transfer through a window. For each trial, they were asked to attribute a mistaken belief about a belief to a character (e.g., "Where will Tom think that Jane will look for her book?") as well as reality (e.g., "Where is the book really?") and memory (e.g., "Where did Jane put the book in the beginning?") control questions.

Children were given a score of '1' for each correctly answered false-belief test question. At both time-points, children were asked 14 test questions: 6 points from the first-order unexpected-contents task (3 for recalling their own belief and 3 for recalling another's false belief), 6 points from the first-order unexpected-transfer task and 2 points from the second-order unexpected-transfer task. All children correctly answered control questions, with the exception of one child with autism who failed the control questions on one second-order trial at Time 2. To avoid loss of data, this child was given a score of 0 out of 2 on this task only.

Executive function. At both time points, children completed two measures each tapping a key component of EF, including planning ability and cognitive flexibility. Both EF components have been identified consistently as being disrupted specifically in autism (see Hill, 2004). The *Tower of London task* (Shallice, 1982; see also Hughes, 1998b) assessed children's higher-order planning ability. Children were presented with 3 colored beads (red, white, black) arranged in a particular configuration (start state) on a wooden tower structure consisting of 3 vertical pegs of increasing size (small, medium, big). They were then shown a picture of the beads in a different configuration (goal state) and asked to move the beads on their apparatus one-at-a-time to match the goal state within the least possible number of moves (this number was clearly stated on the picture). After three practice trials, children were given problem sets of increasing difficulty, including 4 trials of 1-, 2-, 3-, and 4-move problems. At Time 2, an additional set of 5-move problems was included to increase task difficulty and extend the range of possible scores. Testing ceased if children failed all four trials within a problem set. The number of moves taken and rule violations were recorded for each trial. Children were given a score of 1 for each trial if they reached the goal state within the minimum number of moves and without violating any rules (Time 1: maximum score=16; Time 2: maximum score=20). High scores indicate good planning ability.

Cognitive flexibility was assessed using two developmentally-sensitive card-sorting tasks. Both tasks were similar in nature to the traditional WCST (Heaton, 1981), and assessed the ability to switch flexibly between sorting categories in response to feedback. At Time 1, Hughes's (1998a) *teddy-bear set-shifting task* was used. There were three decks of cards, which differed in terms of either color (green vs. pink, blue vs. red, or yellow vs. purple), picture shown (hearts vs. diamonds, squares vs. moons, or stars vs. happy faces), and size of picture (small vs. large). To begin, children were presented with a teddy-bear and one deck of cards, and were told that they were to work out which cards teddy liked best. If the card was one of teddy's favorites, the child posted it into a postbox. Alternatively, if the card was not one of teddy's

favorites, then the child turned the card facedown on the table. Feedback was provided after each trial. When the child had successfully sorted six cards consecutively, or when a maximum of 20 trials had been presented, the sorting rule (e.g., color, shape, size) changed, upon which he/she was presented with a new teddy-bear and new deck of cards. Following Hughes (1998a), children were not alerted to a change in sorting rule; this was implicit in the fact that children were presented with a new situation. Set-shifting performance was rated by the proportion of errors committed following the first sort to criterion.

At Time 2, children were administered a more difficult *computerized set-shifting task* (Comerford, 2005). Children were told that they would see one (target) card that appeared at the top of the screen, and that they would be required to match this card with one of four (response) cards presented at the bottom of the screen. There were 24 response cards, which varied on three dimensions (color, form, and background); all cards shared only one attribute with three out of the four response cards and no attribute with the fourth card. Each card therefore could be sorted according to three rules (color, form, or background). Feedback was provided for each trial: the words 'correct' or 'incorrect' would appear in centre-screen. After the child had sorted 6 consecutive cards correctly, the rule changed (e.g., from color to form) without warning. Testing continued until the set of 24 cards had been presented twice (48 trials in total). To ensure compatibility with the earlier set-shifting task, the main variable of interest was the proportion of errors committed following the first sort to criterion. For both set-shifting tasks, a low score (i.e., minimal errors) indicates good cognitive flexibility.

Central coherence. Two visuospatial tasks were used to assess children's local processing bias: the *Children's Embedded Figures Test* (CEFT) (Witkin, Oltman, Raskin, & Karp, 1971) and the *Pattern Construction task* from the Differential Ability Scales (Elliott, 1990). In the CEFT, children initially were shown a target shape (a triangle) and asked to find this shape hidden in a number of larger meaningful figures as quickly as possible (total 11 trials). There was a single target shape per figure, and children scored one point for each correctly identified target (maximum

score=11). The time taken to find the target also was recorded, and was the dependent variable of interest. If children were unable to locate the target within 30 seconds, the maximum time (30 seconds) was given on that trial. Faster times are indicative of weak CC.

In the Pattern Construction task, which is similar in nature to Wechsler's (1999) Block Design subtest, children were asked to produce a number of increasingly complex patterns using three-dimensional blocks to match a two-dimensional design, as quickly as possible. Similar to the standard Block Design subtest, individual items were scored from a minimum of 0 points (failure) to a maximum (which varied across items), based on accurate reproduction and extra points for speed. Item scores were summed and then converted to ability scores using tables in the Manual (Elliott, 1990), which takes into account the child's raw score and the difficulty of the items administered. Higher scores are indicative of better performance (i.e., weak CC).

Individuals with autism have been shown to outperform comparison individuals on both tasks, and their superior performance has been attributed to an enhanced ability to pre-segment the design into its constituent parts, which enables them to locate rapidly either the hidden figure (in the case of the CEFT) or the appropriate blocks from which to re-construct the pattern (in the case of the Pattern Construction task) (see Happé & Frith, 2006, for review).

General procedure

Ethical approval for the initial and follow-up studies was granted by the Human Research Ethics Committee at the University of Western Australia, and informed written consent was obtained from parents of all children prior to participation. At each time point, children were seen individually on two occasions, each lasting approximately 1-1.5 hours, either at the family home or at the University. Tests of verbal and nonverbal ability were always administered first followed by tests of specific cognitive skills, the order of which was randomized across participants.

Results

The first section presents preliminary descriptions of data screening and analyses on group matching variables. The subsequent sections present the main results for each cognitive domain separately, including descriptive statistics for the individual tasks, and results of analyses assessing developmental persistence and developmental change – two of the primary goals of this study. Where possible, repeated-measures ANOVAs on individual task scores, with "group" (ASD, typical) as the between-participants factor and "time" as the within-participants factor, were used to examine between-group differences and within-group change. The results of hierarchical regression analyses on the ASD data specifically are also reported, which served to identify potential predictors of change in specific cognitive skills. Aggregate scores for each cognitive domain were used as dependent variables. Also, to minimize the number of predictors in these analyses, only those early developmental variables (age, verbal ability, nonverbal ability) that were significantly correlated with the dependent variable were entered as potential predictors. The final section reports the results of analyses to determine whether the putative cognitive profile is characteristic of individual children with autism.

Preliminary data screening. To begin, scores on the set-shifting tasks and the CEFT were reversed so that high scores reflected good performance. Subsequent data screening revealed that the distributions of cognitive variables met assumptions of normality, with the exception of scores on individual ToM tasks, which showed significant positive skew in the scores of children with ASD at each time-point. Transformations were applied to these ToM variables to try to normalize the data without success. Given the robustness of ANOVA against violations of normality, group differences on individual ToM tasks were analyzed using parametric tests. Also, and not unexpectedly, typically developing children performed at ceiling on all tests of false-belief understanding at Time 2. Analyses concerning development changes in ToM skills therefore were restricted to the ASD group. Data screening for outliers (± 3.5 *SD*s of the total sample mean) identified one extreme score on the Pattern Construction task at Time 2 (a child with autism). In line with Wilcox (2002), this score was trimmed by replacing it with the value

representing 3.5 *SD*s above the overall mean, and subsequent analyses were conducted using the trimmed score for this task (note that Table 2 displays the untrimmed scores).

Estimates of reliability were ascertained where possible. For ToM, internal consistency across all 14 trials of the false-belief tasks was high for both children with ASD (Time 1: Cronbach's $\alpha = .87$; Time 2: $\alpha = .94$) and typical children (Time 1: $\alpha = .88$). Cronbach's α could not be calculated for typical children's Time 2 ToM scores since performance was at ceiling for this group. The CEFT (response time) showed modest reliability at both time points in the autism (Time 1: $\alpha = .50$; Time 2: $\alpha = .51$) and typical (Time 1: $\alpha = .51$; Time 2: $\alpha = .47$) groups. It was not feasible to calculate reliability estimates for the Pattern Construction task and EF tasks, as most of these measures incorporate stopping rules as part of their administration. High reliability has been reported, however, for the Pattern Construction task in preschool and school-age children (α ranges between .82 and .90; Elliott, 1990), and for set-shifting tasks in autism (Ozonoff, 1995). For each group, there were significant intercorrelations between individual task scores within each cognitive domain at Time 1 (ToM: $r_{mean} = .66$; EF: $r_{mean} = .45$; weak CC: $r_{mean} = .42$, all ps < .05) and at Time 2 (ToM: $r_{mean} = .55$; EF: $r_{mean} = .63$; weak CC: $r_{mean} = .45$, all ps < .05), indicating good convergent validity. The majority of these correlations remained significant when the potentially confounding effects of verbal ability and nonverbal ability were partialled out (all r_{mean}s > .31).

Group matching. Descriptive statistics for all matching variables are provided in Table 1. There were no significant differences between the ASD and typical groups in terms of chronological age, verbal IQ or nonverbal IQ at either time point (all ps > .12).

Theory of mind

Table 2 shows mean scores for individual tasks at each time point. As expected, at Time 1, children with ASD obtained significantly lower scores compared with typically developing children on the first-order unexpected-contents, $F(1, 67) = 47.70$, $p < .001$, $\eta_p^2 = .42$, and unexpected-transfer, $F(1, 67) = 9.84$, $p < .005$, $\eta_p^2 = .13$, tasks, and the second-order

unexpected-transfer task, $F(1, 67) = 12.06$, $p < .001$, $\eta_p^2 = .15$. This pattern of group differences was also evident at Time 2: autistic children performed significantly worse on both first-order false-belief tasks (unexpected-contents task: $F(1, 67) = 7.02$, $p < .01$, $\eta_p^2 = .10$; unexpected-transfer task, $F(1, 67) = 31.23$, $p < .001$, $\eta_p^2 = .32$), and on the second-order task, $F(1, 67) = 33.33$, $p < .001$, $\eta_p^2 = .34$, than typical children.

Robust aggregate ToM scores for each time point were computed by averaging children's scores from the three individual tasks (see Table 2). A repeated-measures ANOVA on aggregate scores showed main effects of group, $F(1, 66) = 42.08$, $p < .001$, $\eta_p^2 = .39$, confirming the results on individual task scores. There was also a significant main effect of time, $F(1, 66) = 108.40$, $p < .001$, $\eta_p^2 = .62$, indicating a significant *improvement* in ToM scores over the 3-year period across groups. The interaction between time and group was not significant, $F < 1$, although this is most likely attributable to the ceiling effect present in the typically developing data. To ensure that the way in which the reported aggregate score was computed did not mask possible deficits on 2nd-order ToM tasks, two alternative ToM aggregate scores were constructed. One aggregate score was created by standardizing the scores from each of the three ToM tasks and then deriving the average, and a second score was created by averaging the standard scores of the 2nd-order task and the mean of the two 1st-order tasks. Subsequent analyses were carried out using these alternative ToM aggregate scores yet the overall pattern of results remained unchanged.

Insert Table 2 about here

To examine developmental change within individual children with ASD, the number of children passing and failing each ToM task at Time 1 and Time 2 is presented in Table 3. Success was defined conservatively as ≥ 5 out of 6 trials correct (≥ 83%) for each first-order task, and 2 out of 2 trials correct (100%) for the second-order task. McNemar's tests showed a significant increase over time in the number of children with ASD passing the first-order

unexpected-contents task (16% at Time 1 and 78% at Time 2, $p < .001$), the first-order unexpected-transfer task (8% at Time 1 and 49% at Time 2), and the second-order unexpected-transfer task (0% at Time 1 and 43% at Time 2), all ps < .001. All children who passed the first-order ToM tasks at Time 1 also passed these same tasks at Time 2. Of the 9 children who passed either first-order false-belief task at Time 1, all of these children went on to pass the second-order false-belief task at Time 2.

Insert Table 3 about here

A regression analysis using aggregate ToM scores was performed to determine the early predictors of later ToM skills in the ASD group. Time 1 aggregate ToM scores significantly predicted aggregate scores at Time 2 (ß = .48, $R^2 = .23$), $F(1, 36) = 10.57$, $p = .003$. Early verbal ability, $r(35) = .46$, $p < .005$), but not nonverbal ability or chronological age (both ps > .11), was significantly associated with later ToM skills. Early receptive-language scores therefore were entered into the model at the second step of the analysis to determine the additional influence, if any, of this variable on children's later ToM scores. In line with a previous longitudinal study (Steele et al., 2003), Time 1 verbal ability explained variance in Time 2 ToM scores over and above Time 1 ToM (ß = .30, $\Delta R^2 = .08$), $\Delta F(1, 34) = 3.88$, $p < .05$, with better vocabulary skills early on predictive of more advanced ToM skills 3 years later. Early ToM scores remained significantly predictive of children's later ToM scores once variation in early verbal ability had been accounted for (ß = .36, $p < .05$).

Executive function

Children's performance on the Tower of London and Set-shifting tasks is shown in Table 2. For children's Tower of London scores, a repeated-measures ANOVA revealed a significant main effect of group, $F(1, 65) = 37.06$, $p < .001$, $\eta_p^2 = .36$, and time, $F(1, 65) = 125.94$, $p < .001$, $\eta_p^2 = .66$, and a significant group x time interaction, $F(1, 65) = 4.69$, $p < .05$, $\eta_p^2 = .07$. Planned comparisons showed that children with ASD obtained significantly lower

planning scores than typical children both at Time 1, $F(1, 67) = 40.48$, $p < .001$, $\eta_p^2 = .38$, and at Time 2, $F(1, 67) = 14.55$, $p < .001$, $\eta_p^2 = .18$. Analyses further revealed that children's scores improved significantly over time in both groups (ASD: $t(35) = 10.80$, $p < .001$; typical: $t(30) = 5.64$, $p < .001$). The source of the interaction lay instead with the *extent* of improvements over the 3-year period. Children with ASD made significantly *more* gains (M gain = 5.06, $SD = 2.81$) than typical children ($M = 3.42$, $SD = 3.37$), $F(1, 67) = 4.69$, $p < .05$, $\eta_p^2 = .07$, on this measure of higher-order planning.

How many children showed developmental improvements on the Tower of London task? This question was tackled by determining how many children with ASD showed improvements above and beyond what would be expected given their age and ability. The typically developing group made an average gain of 3.4 points on the Tower of London task during the 3-year follow-up period. The majority of children in the ASD group ($n = 26$; 70% of sample) exceeded this level of improvement (average gain of 5.0 points), compared to 12 typically developing children (39%).

Separate one-way ANOVAs were carried out on children's set-shifting performance given the use of different tasks at different time points. Children with ASD performed significantly worse (i.e., made more errors) than typically developing children on the teddy-bear set-shifting task used at Time 1, $F(1, 67) = 18.14$, $p < .001$, $\eta_p^2 = .22$, and on the computerized set-shifting task administered at Time 2, $F(1, 67) = 11.60$, $p < .001$, $\eta_p^2 = .15$, suggesting persistent difficulties with cognitive flexibility.

The significant intercorrelations between Tower of London and Set-shifting task scores meant that an EF aggregate score could be computed at each time point by averaging the standard scores for these tasks. Multiple regression analyses demonstrated that individual differences in early EF aggregate scores strongly predicted ASD children's EF performance at follow-up (ß = .70, $R^2 = .49$), $F(1, 35) = 33.22$, $p < .001$. Cross-time correlations revealed that all developmental variables at Time 1 significantly correlated with EF aggregate scores at Time 2

(age: $r(35) = .30$, $p < .05$; verbal ability: $r(35) = .48$, $p < .005$; nonverbal ability: $r(35) = .52$, $p < .001$), and therefore all were entered into the model at Step 2. Time 1 EF aggregate scores continued to be significant in this model (ß = .66, $p <. 005$) but none of the remaining predictors, however, contributed any unique variance to children's later EF skills (all ps > .42).

Central coherence

Table 2 shows children's CEFT and Pattern Construction scores. An ANOVA on children's CEFT times revealed significant main effects of group, $F(1, 66) = 124.02$, $p < .001$, $\eta_p^2 = .65$, time, $F(1, 66) = 36.09$, $p < .001$, $\eta_p^2 = .35$, and a significant interaction between group and time, $F(1, 66) = 44.84$, $p < .001$, $\eta_p^2 = .40$. As expected, children with ASD showed a significant advantage on this task, with faster times than typically developing children both at Time 1, $F(1, 67) = 168.29$, $p < .001$, $\eta_p^2 = .72$, and Time 2, $F(1, 67) = 10.34$, $p < .002$, $\eta_p^2 = .14$. For typically developing children, the time taken to find the hidden figure reduced significantly over the 3-year period, $t(30) = 7.49$, $p < .001$. This was not the case, however, for children with ASD, whose times remained unchanged, $t < 1$. Comparable results were obtained when median times were used in analyses.

An ANOVA on children's Pattern Construction scores revealed main effects of group, $F(1, 66) = 20.75$, $p < .001$, $\eta_p^2 = .24$, time, $F(1, 66) = 60.52$, $p < .001$, $\eta_p^2 = .48$, and a significant group x time interaction, $F(1, 66) = 30.87$, $p < .001$ $\eta_p^2 = .32$. Planned comparisons showed that children with ASD performed significantly better than comparison children at Time 1, $F(1, 67) = 50.11$, $p < .001$, $\eta_p^2 = .43$, but failed to maintain their advantage on this task at Time 2, $F(1, 67) = 1.45$, $p = .23$, $\eta_p^2 = .02$. Further within-group analyses showed that, similar to the pattern of results obtained for the CEFT, typically developing children's scores improved significantly over time, $F(1, 30) = 118.68$, $p < .001$, $\eta^2 = .80$, but ASD children's scores did not, $F(1, 36) = 2.15$, $p = .15$, $\eta^2 = .06$. Examination of individual children's performance over time was addressed in the same way as it had been for children's Tower of London performance. While the ASD group improved, on average, only 4.8 points on the Pattern Construction task between

intake and follow up, typically developing children improved, on average, 25.8 points. Twelve typical children (39%) demonstrated this level of improvement compared with no child in the ASD group.

CC aggregate scores for each time point were created by standardizing the CEFT time and Pattern Construction scores, and averaging them. Multiple regression analyses revealed that CC aggregate scores at intake were significantly predictive of ASD children's scores 3 years later, (ß = .56, R^2 = .31), $F(1, 35) = 15.85$, $p < .001$. Since Time 2 CC aggregate scores were significantly correlated with initial verbal ability, $r(35) = .64$, $p < .001$, and nonverbal ability, $r(35) = .67$, $p < .001$, but not with age (p=.10), the former two Time 1 variables were added to the model at the second step. Nonverbal ability alone contributed an additional 19% of the variance in later CC aggregate scores, (ß = .50, ΔR^2 = .19), $\Delta F(1, 34) = 12.66$, $p < .001$, while initial CC scores remained significant (ß = .32, $p < .05$), once variation in verbal and nonverbal ability had been accounted for.

The fact that the typically developing group improved considerably on both CC measures might imply, paradoxically, that children develop a more pronounced local processing bias over time. Other factors, such as children's maturing nonverbal and executive skills, however, are most likely driving this developmental change. Indeed, performance on both CC measures has been linked previously to concurrent verbal and nonverbal ability (Wechsler, 1999; Witkin et al., 1971), and early nonverbal ability (but not verbal ability) was longitudinally predictive of later performance on the Pattern Construction task, $r(29) = .51$, $p < .005$, and the CEFT, $r(29) = .33$, $p = .06$, in the current sample of typical children. Analyses also revealed a trend for early EF aggregate scores to be associated with later Pattern Construction performance, $r(29) = .32$, $p = .07$, and CEFT performance, $r(29) = .30$, $p = .09$, consistent with the possibility that typically developing children's improving executive skills might enable children to better resist interference from the overall Gestalt and therefore rapidly locate either the hidden figure (in the case of the CEFT) or the blocks necessary to reconstruct the design (in

the case of the Pattern Construction task). Since the scores on both CC measures do not represent "pure" indices of local processing, the two CC variables (children's CEFT and Pattern Construction scores) were regressed on nonverbal ability and EF at each time point, and the unstandardized residuals for each variable were saved. For each individual, these residual scores represent the difference between his/her observed score and what would be expected given his/her nonverbal and executive ability. The saved residuals for each variable were then converted to standard scores, and averaged to form an aggregate CC score for each time point. These aggregate scores were used in subsequent analyses to address the question of universality of enhanced local processing skills in autism.

Multiple cognitive atypicalities at the individual level

The results at the group level demonstrate that children with ASD do in fact display the cognitive profile proposed by Happé et al. (2006b), and that in general this profile persists across time. The next set of analyses examined whether *individual* children with ASD exhibited this profile at intake and follow-up.

To examine the universality of cognitive atypicalities, the number of children with ASD that displayed an "atypicality" was calculated for each aspect of cognition using the modified t test developed by Crawford and Howell (1998). This one-tailed test is well-suited for comparing single cases to small comparison groups since it treats the statistics of the comparison sample as statistics rather than as population parameters, and further provides excellent control over Type I error rates. These authors use Sokal and Rohlf's (1995) formula,

$$t = \frac{X_1 - X_2}{S_2\sqrt{(N_2 + 1)/N_2}},$$

where X_1 is an individual ASD child's score, X_2 and S_2 are the mean and standard deviation of scores from the typically developing group, and N_2 is the comparison group sample size. Following Crawford and Howell (1998), the *p* value generated by each test is taken to be an estimate of the atypicality of an individual's score. For example, a *p* value of .15 indicates the

proportion of the comparison group who would receive a score of similar or greater magnitude than the score obtained for a particular individual. For the purposes of this study, an individual's aggregate score for each of the three domains was classified as "atypical" if the *p* value fell below .15. Note that this definition is functionally similar to previous definitions of an atypicality (Lezak, 1991; see Pellicano, 2007) in which a score is considered to be atypical if it exceeds 1 standard deviation above/below the mean score of the typically developing group. Venn diagrams illustrating the results can be found in Figure 1.

Insert Figure 1 about here

At Time 1, 27 out of 37 children with ASD (73% of the group) scored below the cut-off ($p < .15$) on the ToM aggregate, 24 children (65%) showed an impairment in EF, and remarkably, all but one child (97%) displayed weak CC. How many children showed atypicalities in *all three cognitive domains* at intake (cf. Happé et al., 2006b)? More than half of the ASD group (59%) showed co-occurring atypicalities in ToM, EF, and CC. Five children (14%) showed poor ToM and weak CC, two children (5%) showed atypicalities in both EF and CC, and the remaining 7 children (19%) displayed weak CC alone (see Figure 1a). One additional child (3%) showed no atypicalities at Time 1.

At Time 2, 26 out of 37 children with ASD (70%) fell below the cutoff on the ToM aggregate, 20 children (54%) showed poor EF and 18 children (49%) showed enhanced local processing. Only 7 children (19%) showed multiple atypicalities in ToM, EF and CC at Time 2 (see Figure 1b). Of the remaining 30 children, a substantial portion (11 children: 30%) showed joint problems in ToM and EF, 6 children (16%) displayed weak CC combined with poor ToM, 1 child (3%) showed weak CC and poor EF, 2 children (5%) showed ToM difficulties in isolation, 4 children (11%) showed only weak CC, and 1 child (3%) showed EF difficulties alone. Surprisingly, 5 children (13%) showed *no* cognitive atypicalities at follow-up. Of these 5

children, 4 had a clinical diagnosis of autism and 1 child had a diagnosis of PDD-NOS, and 4 of the 5 children showed sufficient behavioral symptoms at follow-up to meet the cut-off for autism on the ADOS-G (Lord et al., 1999). Among the 22 children who showed all three atypicalities at Time 1, 5 children continued to display this profile at Time 2, 11 children showed persistent difficulties in ToM and EF, 2 children showed atypicalities in ToM and CC, 3 children showed ToM problems in isolation and 1 child showed no atypicalities.

It is possible that, among those children who failed to demonstrate atypicalities in one or more cognitive domains, such atypicalities were in fact present in some or all of these children, albeit to a lesser degree. To investigate this further, additional analyses were carried out using a more lenient definition of atypicality, where an individual's aggregate score for each cognitive domain was identified as "atypical" if the p value fell below .50. (Note that this is a lenient definition of atypicality since the p value estimates the point at which 50% of the typical population would obtain scores lower than the score obtained for a particular individual, cf. Crawford & Howell, 1998.) According to the main universality analyses using the conservative criterion ($p < .15$; see Figure 1), at Time 1, 10 children showed no problems in ToM, 13 children demonstrated no EF difficulties, and 1 child failed to show weak CC. Subsequent analyses using the more lenient criterion ($p < .50$) revealed that all of these cases showed atypicalities in these domains (ToM: all cases $\leq p = .48$; EF: all cases $\leq p = .43$; CC: all cases $\leq p = .43$). At Time 2, the central, more conservative analyses revealed that 11 children showed no ToM difficulties, 17 children showed no EF problems, and 19 children failed to show weak CC. Again, when the more lenient criterion was applied, all children were classified as atypical on each aspect of cognition (ToM: all cases $\leq p = .38$; EF: all cases $\leq p = .46$; CC: all cases $\leq p = .45$), suggesting that this entire sample of children with ASD appeared to show the putative cognitive profile, albeit in subtle form.

Discussion

Developmental persistence

Relative to age- and ability-matched typically developing children, children with ASD showed difficulties in false-belief understanding, problems with higher-order planning and cognitive flexibility together with capabilities in processing local information *at both time points*. This finding supports Happé et al.'s (2006b) claim that there is a specific profile of coexisting cognitive atypicalities in autism, which shows continuity with development. Yet testing the claims of Happé et al.'s multiple-deficits model required not only analysis of the group as a whole but also examination of the performance of individual children with ASD. Such analysis demonstrated that the putative profile was not universal at either time point, and became markedly less pervasive at follow-up. At intake, weaknesses in false-belief attribution and aspects of EF were found to be less pervasive than atypicalities in CC, resulting in just over half of the ASD group displaying coexisting atypicalities in all three cognitive domains. At follow-up, a dramatically different pattern emerged: none of the atypicalities were entirely pervasive, and only a minority of children displayed the complete profile. Inspection of the group means at Time 2 revealed that the distributions overlapped considerably, but particularly so for the weak CC tasks. Indeed, group differences on the CEFT were attenuated at Time 2, and had disappeared all together on the Pattern Construction task.

Measurement issues might have contributed to the failure to demonstrate the universality of this cognitive profile in two ways. First, the Time 2 analyses on ToM were complicated by the presence of ceiling effects in the typically developing group, which could have influenced the magnitude of the group difference, and in turn, the number of individual children with ASD showing ToM difficulties. Even in the event that this was true, it is difficult to see how this explanation alone could explain the minority of children with ASD (n=7) showing all three atypicalities at follow-up since this ceiling effect should not have affected the results concerning atypicalities in EF and CC, which together accounted for less than half of the group.

Second, the cognitive measures were matched neither for discriminative power nor for reliability (cf. Chapman & Chapman, 1973) rendering it possible that the tasks were not sufficiently sensitive to detect subtle cognitive atypicalities. Internal consistency was moderate-to-high for cognitive tasks, which remained stable across time points, and there was good convergent validity for each cognitive domain at intake and follow-up. Reliability was not evaluated for executive measures, which has been reported to be low in typical samples (e.g., Bishop et al., 2001) and high in autism samples (e.g., Ozonoff, 1995). The results of these analyses therefore should be treated with some caution.

Despite these limitations, the fact that the majority of children failed to show multiple cognitive atypicalities at follow-up combined with the presence of some children who showed *no* cognitive atypicalities is problematic for Happé et al.'s (2006b) candidate multiple-deficits account, which attempts to provide an integrated causal explanation for the pathogenesis of autism. It is noteworthy that analyses using a more lenient definition of atypicality demonstrated that *all* children with ASD showed the putative cognitive profile, which is encouraging for Happé et al.'s account. It is worth emphasizing, however, that according to this lenient definition of atypicality ($p<.50$), one should expect to find 50% of individuals in the population from which the comparison sample was drawn also to receive a score of similar or greater magnitude than that observed for individuals with ASD (Crawford & Howell, 1998). In such instances, one would be cautious in classifying these typical children as showing an atypicality, and therefore one must be equally cautious of doing so in the case of ASD.

Certainly, the findings from these supplementary analyses serve to highlight the challenges in defining atypicality, especially when examining individuals with developmental conditions, whose cognitive skills are unlikely to be "all or none". Rather than 'atypicality' being conceptualized categorically, it might be beneficial to conceive of the three cognitive domains as dimensions located (orthogonally) within a multivariate space (Happé et al., 2006b; Happé & Ronald, 2008). The extent to which a person with autism shows a particular cognitive atypicality

therefore should vary according to the place he/she occupies on that dimension, which in turn, should relate directly to the degree and nature of the behavioral symptoms it purports to explain. This elaboration is appealing as it could potentially account for the phenotypic heterogeneity in autism. It would be of interest, therefore, to demonstrate empirically that individual differences in cognitive skills are indeed significantly related to individual differences in specific aspects of the behavioral phenotype – both concurrently and longitudinally. Despite its appeal, it nevertheless remains a challenge for this account to explain the absence of any of the suggested cognitive atypicalities in 4 children with ASD who have sufficient behavioral symptoms at follow-up. It is perhaps plausible that the emergence of cognitive atypicalities at some point in development could be sufficient to cause behaviors severe enough to persist beyond the cognitive atypicalities themselves.

Developmental progress in autistic children's cognitive skills

An understanding of such putative processes necessarily demands a fuller developmental perspective than is currently offered. The second main finding from this study was that were considerable developmental improvements in autistic children's ToM and EF, but no changes in local processing. Children with ASD made substantial progress in their false-belief understanding over the 3-year period to the extent that more than one third of the group succeeded on advanced false-belief tasks at follow-up. These findings present a more optimistic picture of autistic children's developing ToM than earlier longitudinal studies, which reported negligible progress in children's false-belief understanding (Holroyd & Baron-Cohen, 1993; Ozonoff & McEvoy, 1994). Notably, children in these early investigations were older and less able than participants sampled both in the current study and in two recent longitudinal studies, which also demonstrated improvements in some aspects of children's ToM skills (Serra et al., 2002; Steele et al., 2003), highlighting the possibility that improvements, should they occur, might be most apparent early on during development (cf. Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998). The greater emphasis in recent years on early behavioral intervention for young

children with autism might also account for the greater developmental progress reported in more recent studies compared to earlier ones.

Children's planning capacity also improved significantly over the 3-year period – at a strikingly faster rate than that of typically developing children. These findings go against the two existing prospective studies on EF in autism, which reported progress in EF neither in very young children with autism (Griffith et al., 1999) nor in adolescents with autism (Ozonoff & McEvoy, 1994). Again, differences in sample selection could explain the opposing pattern of results since both of these studies included individuals with autism less able than the sample of autistic children assessed here. Interestingly, and akin to the current findings, Happé, Booth, Charlton, and Hughes (2006a) found that the EF performance of older participants with autism (M age=13 years 2 months) surpassed that of younger children (M age=9 years 2 months), suggestive of age-related gains in EF. Furthermore, although the authors did not find evidence of specific problems in planning ability, they nonetheless showed that this EF component showed the greatest age-related improvements in the autism group. The current longitudinal data strengthen Happé et al.'s (2006a) cross-sectional findings, and indicate that EF problems become *less marked with age*, at least for cognitively-able children with autism.

Several factors might explain this boost in ASD children's ToM and executive skills over the 3-year-period. The first of these is language. Regression analyses showed that verbal ability contributed significant variance in later ToM scores over and above variance already accounted for by early ToM performance. The important influence of language during children's developing ToM is consistent with an abundance of cross-sectional (e.g., Fisher, Happé, & Dunn, 2005; Happé, 1995; Leekam & Perner, 1991) and some longitudinal (Steele et al., 2003; Tager-Flusberg & Joseph, 2005) work, showing that language ability, and especially grammatical ability, is significantly related to autistic children's performance on ToM tasks. Alongside these findings, the current data provide firm evidence that early language skills might play an

important role in facilitating autistic children's understanding of the representational nature of mind.

Early verbal ability was, however, unrelated to children's developmental improvements in planning ability. This finding was unexpected since there are both empirical and theoretical reasons to anticipate a link between language and children's developing executive skills. Russell (1996) argued that (internal) language ability plays a key role in executive control over action. Specifically, success on executive tasks, such as the Tower of London task, should be bolstered by the use of inner speech, i.e., if the child forms a *verbal representation* of the sequence of moves necessary to solve the problem in the minimum number of moves. Two studies have shown experimentally that reduced executive control in autism might indeed result from an inherent failure to use language to regulate and control one's actions (Joseph, Steele, Meyer, & Tager-Flusberg, 2005; Whitehouse, Maybery, & Durkin, 2006). If language is integral to success on EF tasks, then it is puzzling that early language skills were unrelated to developmental *improvements* on such tasks in the current study. Perhaps children's use of compensatory strategies, which might be non-verbal in nature, becomes increasingly practiced over time, affording gradual gains in executive control.

The second factor that could explain the enrichment in autistic children's developing ToM and EF is social contact (Hughes, 1998b). Work with typically developing children shows that early peer interactions, particularly those involving pretence, foster the development of ToM (Hughes & Dunn, 1998; see Dunn, 2004, for review). Likewise, social relations are likely to influence children's developing executive skills since negotiating everyday social interactions requires children to regulate their own behaviors (Luria, 1966, see also Hughes, 1998b). The majority of autistic children in the current study attended mainstream classrooms (n=34), and therefore encountered frequent (and varying) models of age-appropriate social interactions. Evidence suggests that cognitively-able children with autism seek out their non-autistic peers (Sigman & Ruskin, 1999). Engaging with more advanced social partners on an everyday basis

therefore might have bootstrapped the development of these children's mental-state reasoning and executive control (Vygotsky, 1978). Indeed, increased social contact during this period of development might also account for the *enhanced* rate of acquisition of EF skills compared to typical children. EF shows a protracted development trajectory, which renders it especially susceptible to environmental influences. Recent findings suggest that EF skills are indeed malleable in young children (Diamond, Barnett, Thomas, & Munro, 2007). It therefore remains possible that attendance in mainstream schools provided an enriched social environment with ample opportunities for autistic children to "exercise" their EF skills, therefore boosting the rate of acquisition of such skills. Identifying which aspects of social relations, if any, partially mediate the development of children's EF and ToM skills is a worthy line of investigation since this should have important implications for the remediation of atypicalities in these cognitive domains.

A third reason for the improvements in EF and ToM is that the development of these functions is in fact inextricably linked. There is strong evidence that ToM and EF are closely tied in typically developing children (e.g., Carlson & Moses, 2001; Hughes, 1998a, b) and in autism (e.g., Pellicano, 2007). The nature of this link, however, is controversial. Some authors have proposed that executive skills are a prerequisite for the later development of ToM (Moses, 2001; Russell, 1996), while others claim that a representational understanding of mind is necessary for the later control of goal-directed, purposeful action (Perner & Lang, 1999). Further still, the parallel progression of ToM and EF might be due to a third factor: that both skills are mediated by adjacent structures in the pre-frontal cortex (Bishop, 1993; Ozonoff et al., 1991). Research on the overlap between these two domains in autism is scarce yet analysis of the longitudinal links between EF and ToM in the current sample of children with ASD (see Pellicano, 2009) points towards a developmental relationship in one direction only. Earlier EF skills were found to be longitudinally predictive of autistic children's later ToM test performance, independent of age, verbal ability, nonverbal ability, and early ToM skills but there

were no significant predictive relations in the opposite direction. Akin to work with typically developing children, these findings find favor with Russell's account that EF plays an important role in the advancement of autistic children's ToM skills (see also Pellicano, 2007). Nevertheless, the possibility of a functional link between EF and ToM is potentially inconsistent with Happé et al.'s (2006b) model, which considers the co-occurring cognitive atypicalities in autism to be independent and genetically distinct, and is therefore unlikely to support any notion of one atypicality emerging from another.

Despite significant improvements in ToM and executive skills, children experienced no changes in local processing over the 3-year period. Children with ASD were no quicker to find the hidden figure on the CEFT than they were 3 years earlier, and their scores on the Pattern Construction task failed to change over time. This lack of change is in striking contrast to the significant gains made by typically developing children evidenced on both weak CC tasks. Could the lack of an improvement be explained by a ceiling effect in the ASD group? This explanation might account for children's CEFT performance, but it is unlikely that a ceiling effect could account for children's equally limited improvement on the Pattern Construction task since the majority of children's scores were well below the maximum score.

One alternative explanation is that the development of local processing takes strikingly different courses in autism and typical development. This suggestion is consistent with one competing account of perceptual atypicalities in autism, enhanced perceptual functioning (Mottron & Burack, 2001), which proposed that such peaks derive from the overdevelopment of basic perceptual processes, including local processing. Rather than being present from birth, superior local processing emerges early on during the course of development in response to diminished processing of higher-order operations. In line with Mottron and Burack's (2001) claim, the current findings suggest these skills are early-emerging and initially accelerated in autism. Over time, it appears that the trajectories of children with ASD and typically developing

children converge as the local processing skills of typical children "catch up" to that of autistic children.

The current data also highlight the possibility that the developmental trajectory of local processing in autism might be qualitatively distinct from the trajectories of other, higher cognitive domains (ToM, EF), which appear to mature later, and progress somewhat in parallel. Do these findings fit with Happé et al.'s (2006b) theoretical position? In its current form, the model says little about the ways in which these cognitive atypicalities unfold with development, and which sort of factors influences their progression. The possibility of distinct developmental trajectories for some aspects of cognition might be taken as evidence to support a fractionable cognitive characteristics account (Happé & Ronald, 2008). Although Happé et al. (2006b) hinted at the possibility of potential "interactions" between cognitive domains, the nature of any such interactions remains unspecified. Analysis of the current longitudinal data (see Pellicano, 2009) suggest that interactions do in fact exist, specifically in which early domain-general skills (EF and CC) play a crucial role in shaping the developmental trajectory of autistic children's emerging ToM. Certainly, further elaboration of Happé et al.'s model is required to account for the continuities and discontinuities in the development of core cognitive skills, the specific *developmental* relationships between cognitive domains, in addition to the potential factors (e.g., verbal ability, social interaction; see Hughes, 1998b) that might mediate or moderate such development.

Conclusion

This is the first prospective study to investigate the development of *multiple* cognitive atypicalities in ASD. Several tasks tapping ToM, EF, and weak CC, were administered to the same samples of children at two different time points, separated by three years. The findings showed that cognitively-able children with ASD, as a group, do show coexisting atypicalities in important aspects of ToM, EF, and weak CC, relatively to their typically developing comparison children, and that this profile persists over time – precisely in the way that Happé et al. (2006b)

proposed. Their multiple-deficits model goes far beyond previous single-deficit models by invoking several core underlying atypicalities that, together, cause autism – arguably, a much more realistic position given the heterogeneity present in the condition. Yet not all children showed this particular cognitive profile. Nor did children's cognitive skills remain static and unchanging. In fact, they experienced considerable gains in those aspects of cognition which typically present them with the most difficulty – progress which is encouraging and perhaps unsurprising in developing cognitive systems, but is nonetheless difficult to predict from Happé et al.'s model in its current form. Whether this pattern of findings will be similar for children with autism who have additional learning difficulties remains a crucial question for future research. The current findings nevertheless stress the need for theorists to take seriously the *developing* cognitive phenotype of autism.

References

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV, 4th Ed.)*. Washington, D.C: Author.

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

Baron-Cohen, S., Tager-Flusberg, H., & Cohen. D. J. (2000). *Understanding other minds: perspectives from developmental cognitive neuroscience.* Oxford, UK: Oxford University Press.

Bishop, D. V. M. (1993). Annotation: Autism, executive functions and theory of mind: A neuropsychological perspective. *Journal of Child Psychology and Psychiatry, 34,* 279–293.

Bishop, D. V. M., Aamodt-Leeper, G., Creswell, C., McGurk, R., & Skuse, D. H. (2001). Individual differences in cognitive planning on the Tower of Hanoi task: Neuropsychological maturity or measurement error? *Journal of Child Psychology and Psychiatry, 42,* 551-556.

Bolton, P., Macdonald, H., Pickles, A., Rios, P., Goode, S., Crowson, M., et al. (1994). A case–control family history study of autism. *Journal of Child Psychology & Psychiatry, 35,* 877–900.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72,* 1032–1053.

Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficit. *Psychological Bulletin, 79,* 380-385.

Comerford, B. (2005). Can inner speech deficits explain executive function deficits in autism? University of Western Australia (unpublished thesis).

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *Clinical Neuropsychologist, 12,* 482-486.

de Villiers, J. (2000). Language and theory of mind: What are the developmental relationships. In S. Baron-Cohen, H. Tager- Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience, Second Edition* (pp. 83–123). Oxford: Oxford University Press.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318,* 1387-1388.

Dunn, J. (2004). *Children's friendships: The beginning of intimacy.* Oxford: Blackwell publishing.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test—Third Edition.* Circle Pines, MN: American Guidance Service.

Elliott, C. D. (1990). *Differential Ability Scales.* New York: Psychological Corporation.

Fisher, N., Happé, F., & Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry, 46,* 409-419.

Frith, U. (1989). *Autism: Explaining the enigma.* Oxford, UK: Blackwell Publishers.

Frith, U. (2003). *Autism: Explaining the enigma* (2nd ed.). Oxford: Blackwell.

Griffith, E. M., Pennington, B. F., Wehner, E. A., & Rogers, S. J. (1999). Executive functions in young children with autism. *Child Development, 70,* 817–832.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66,* 843–855.

Happé, F., Booth, R., Charlton, R., & Hughes, C. (2006a). Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: Examining profiles across domains and ages. *Brain and Cognition, 61,* 25-39.

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36,* 5-25.

Happé, F. & Ronald, A. (2008). The 'fractionable autism triad': A review of evidence from behavioural, genetic, cognitive, and neural research. *Neuropsychology Review, 18,* 287-304.

Happé, F., Ronald, A., & Plomin, R. (2006b). Time to give up on a single explanation of autism. *Nature Neuroscience, 9,* 1218-1220.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test manual.* Odessa, FL: Psychological Assessment Resources.

Hill, E. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences, 8,* 26-32.

Holroyd, S., & Baron-Cohen, S. (1993). Brief report: How far can people with autism go in developing a theory of mind? *Journal of Autism and Developmental Disorders, 23,* 379–386.

Hughes, C. (1998a). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16,* 233–253.

Hughes, C. (1998b). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34,* 1326–1339.

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental-state talk between young friends. *Developmental Psychology*, *34*, 1026–1037.

Hughes, C., & Russell, J. (1993). Autistic children's difficulty with mental disengagement from an object: Its implications for theories of autism. *Developmental Psychology, 29,* 498-510.

Joseph, R. M., Steele, S. D., Meyer, E., & Tager-Flusberg, H. (2005). Self-ordered pointing in children with autism: Failure to use verbal mediation in the service of working memory. *Neuropsychologia, 43,* 1400–1411.

Leekam, S. R., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition, 40,* 203–218.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Liss, M., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., et al. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology and Psychiatry, 42,* 261–270.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). Autism Diagnostic Observation Schedule—Generic. Los Angeles: Western Psychological Services.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview—Revised. *Journal of Autism and Developmental Disorders, 24,* 659–685.

Luria, A. R. (1966). *Higher cortical functions in man*. New York: Basic Books.

Mandy, W. P., & Skuse, D. H. (2008). Research review: what is the association between the social-communication element of autism and repetitive interests, behaviours and activities? *Journal of Child Psychology and Psychiatry, 49,* 795–808.

Morgan, B., Maybery, M., & Durkin, K. (2003). Weak central coherence, poor joint attention, and low verbal IQ: Independent deficits in early autism. *Developmental Psychology*, *39*, 646–656.

Moses, L. J. (2001). Executive accounts of theory of mind development. *Child Development, 3,* 688–690.

Mottron, L., & Burack, J. (2001). Enhanced perceptual functioning in the development of persons with autism. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 131 –148). Hillsdale, NJ: Erlbaum.

Ozonoff, S. (1995). Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. *Neuropsychology, 9,* 491–500.

Ozonoff, S., & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind development in autism. *Development and Psychopathology, 6,* 415–431.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32,* 1081-1105.

Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: clues to developmental primacy. *Developmental Psychology, 43,* 974-990.

Pellicano, E. (2009). Individual differences in executive function and central coherence predict later understanding of mind in autism. Manuscript submitted for publication.

Pellicano, E., Maybery, M., Durkin, K., & Maley, A. (2006). Multiple cognitive capabilities/deficits in children with an autism spectrum disorder: 'Weak' central coherence and its relationship to theory of mind and executive control. *Development and Psychopathology, 18,* 77-98.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37,* 51–87.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3,* 337–344.

Perner, J., Leekam, S., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5,* 125–137.

Perner, J., & Wimmer, H. (1985). "John *thinks* that Mary *thinks* that . . .": Attributions of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39,* 437–471.

Piven, J., Palmer, P., Jacobi, D., Childress, D., & Arndt, S. (1997). Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. *American Journal of Psychiatry, 154,* 185–190.

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review, 27,* 224-260.

Roid, G. H., & Miller, L. J. (1997). *Leiter International Performance Scale—Revised.* Wood Dale, IL: Stoelting.

Ronald, A., Happé, F., Bolton, P., Butcher, L. M., Price, T. S., Wheelwright, S., et al. (2006). Genetic heterogeneity between the three components of the autism spectrum: a twin study. *Journal of the American Academy of Child & Adolescent Psychiatry, 45,* 691–699.

Russell, J. (1996). *Agency: Its role in mental development.* Hove, England: Erlbaum.

Rutter, M. (1983). Cognitive deficits in the pathogenesis of autism. *Journal of Child Psychology and Psychiatry, 24,* 513-531.

Rutter, M., Bailey, A., & Lord, C. (2003). *Social Communication Questionnaire.* Los Angeles, CA: Western Psychological Services.

Serra, M. Loth, F. L., van Geert, P. L. C., Hurkens, E., & Minderaa, R. B. (2002). Theory of mind in children with 'lesser variants' of autism: a longitudinal study. *Journal of Child Psychology and Psychiatry, 43,* 885-900.

Shallice, T. (1982). Specific impairments in planning. *Philosophical Transactions of the Royal Society of London, Series B, 298,* 199–209.

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. *Monographs of the Society for Research in Child Development, 64,* 1–114.

Sokal, R. R., & Rohlf, J. F. (1995). *Biometry*. San Francisco, CA: W.H. Freeman.

Steele, S., Joseph, R. M., & Tager-Flusberg, H. (2003). Developmental change in theory of mind abilities in children with autism. *Journal of Autism and Developmental Disorders, 33,* 461–467.

Tager-Flusberg, H., & Joseph, R. M. (2005). How language facilitates the acquisition of false-belief understanding in children with autism. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 298-318). Oxford, UK: Oxford University Press.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. New York: The Psychological Corporation.

Whitehouse, A. J. O., Maybery, M. T., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry, 47,* 857–865.

Wilcox, R. R. (2002). Understanding the practical advantages of modern ANOVA. *Journal of Clinical Child and Adolescent Psychology, 31,* 399-412.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. (1971). *A manual for the Embedded Figures Test*. Palo Alto, CA: Consulting Psychologists Press.

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities in individuals with autism, individuals with mental retardation and normally developing individuals. *Psychological Bulletin*, *124*, 283–307.

Author Note

Elizabeth Pellicano, Centre for Research in Autism and Education, Department of Psychology and Human Development, Institute of Education, London, and the School of Psychology, University of Western Australia.

This research was supported by the British Academy, the Experimental Psychology Society (UK), and University College, Oxford. The author is grateful to all the families for their continued support, and to Murray Maybery and Marc Stears for helpful discussions.

Correspondence concerning this article should be addressed to Liz Pellicano, Centre for Research in Autism and Education, Department of Psychology and Human Development, Institute of Education, 25 Woburn Square, London, WC1H 0AA, UK. Electronic mail may be sent to l.pellicano@ioe.ac.uk

Table 1. *Descriptive statistics for chronological age, verbal ability, and nonverbal ability for the autism (n=37) and typically developing (n=31) groups at both time points.*

Variable	**Autism** *M (SD)* Range	**Typical** *M (SD)* Range	*F*	*p*
Chronological age				
Time 1	67.92 (10.42) 49 – 88	65.19 (12.64) 48 – 88	.95	.33
Time 2	100.84 (11.15) 80 – 123	97.94 (13.48) 75 – 122	.94	.34
Verbal IQ (PPVT-III)				
Time 1	97.08 (11.52) 80 – 122	100.97 (8.72) 87 – 120	2.38	.13
Time 2	93.89 (17.88) 62 – 138	99.64 (10.53) 80 – 123	2.48	.12
Nonverbal IQ (Leiter-R)				
Time 1	113.27 (13.93) 83 – 141	115.61 (16.42) 89 – 147	.40	.53
Time 2	104.35 (12.72) 80 – 135	106.97 (10.16) 93 – 129	.85	.36

Table 2. *Mean scores for tasks tapping each cognitive domain in the autism (n=37) and typically developing (n=31) groups at intake and follow-up.*

Variable	Group			
	Autism (*n*=37)		**Typically developing (*n*=31)**	
	Time 1 *M (SD)* Range	**Time 2** *M (SD)* Range	**Time 1** *M (SD)* Range	**Time 2** *M (SD)* Range
Theory of mind				
First-order unexpected-contents (out of 6)	1.92 (2.13) 0 – 6	4.97 (1.99) 0 – 6	4.84 (1.10) 3 – 6	5.94 (.36) 4 – 6
First-order unexpected-location (out of 6)	1.78 (1.83) 0 – 6	3.32 (2.62) 0 – 6	3.35 (2.30) 0 – 6	5.97 (.18) 5 – 6
Second-order unexpected-location (out of 2)	.08 (.28) 0 – 1	1.00 (0.94) 0 – 2	.61 (.88) 0 – 2	1.94 (.25) 1 – 2
Aggregate score (maximum score = 4.67)	1.26 (1.25) 0 – 4.33	3.10 (1.58) 0 – 4.67	2.94 (1.30) 1 – 4.33	4.61 (.17) 4 – 4.67
Executive Function				
Tower of London (no. of trials solved in min. no. of moves)	6.33 (2.80)* 2 – 13	11.39 (3.21) 5 – 18	10.52 (2.67)* 4 – 15	13.94 (2.01) 9 – 18
Set-shifting† (prop. errors following first sort to criterion)	.32 (.07) .18 – .45	.44 (.16) .22 – .80	.23 (.09) .11 – .50	.33 (.08) .20 – .48
Central coherence				
CEFT (score out of 11)	9.73 (1.15) 7 – 11	9.73 (1.02) 7 – 11	8.42 (1.06) 7 – 11	9.42 (1.23) 6 – 11
CEFT (s)	5.46 (2.88) 1.51 – 12.29	5.79 (2.77) 1.21 – 12.83	14.11 (2.56) 8.44 – 18.41	8.04 (2.99) 3.42 – 15.31
Pattern Construction	132.14 (17.64) 108 – 182	136.81 (20.90) 104 – 211	105.64 (12.08) 75 – 131	131.42 (13.54) 111 – 173

Notes: * *n*=36

† Different tasks tapping cognitive flexibility were performed at each time point.

Table 3. *Numbers of children with autism (n=37) passing and failing individual theory of mind tasks at intake (Time 1) and follow-up (Time 2).*

	Time 2	
Time 1	**Fail**	**Pass**
First-order unexpected-contents		
Fail: ≤ 4 points)	8	23
Pass: 5 or 6 points	0	6
First-order unexpected-transfer		
Fail: ≤ 4 points	19	15
Pass: 5 or 6 points	0	3
Second-order unexpected-transfer		
Fail: ≤ 1 point	21	16
Pass: ≥ 2 points	0	0

Figure caption

Figure 1. The overlapping circles in these Venn diagrams represent the potential combinations of atypicalities in ToM, EF, and CC. The region in the centre represents co-occurring atypicalities in all three cognitive domains: the putative cognitive phenotype of autism (cf. Happé et al., 2006b). The numbers inside each region represent the number of children with ASD showing that particular atypicality alone or combination of atypicalities at (a) Time 1 and (b) Time 2.

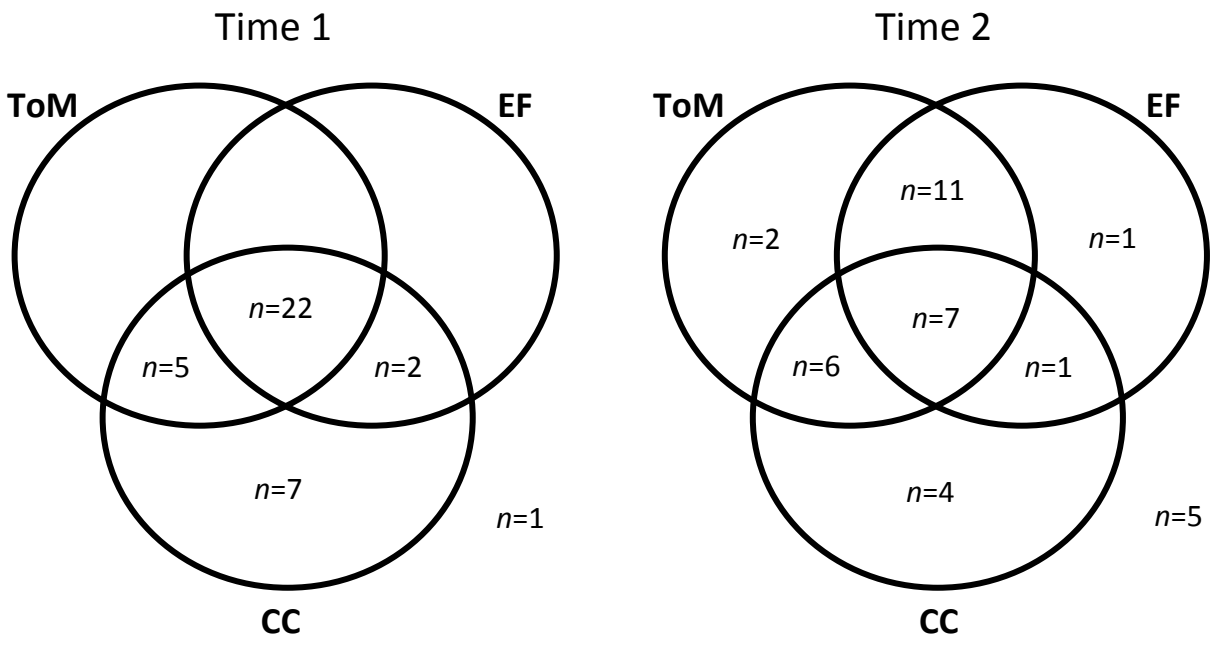
Time 1
ToM
EF
n=22
n=5
n=2
n=7
n=1
CC
(a)
Time 2
ToM
EF
n=11
n=2
n=1
n=7
n=6
n=1
n=4
n=5
CC
(b)

Figure 1.