16 What Neurodevelopmental Disorders Can Reveal about Cognitive Architecture: The Example of Theory of Mind

Helen Tager-Flusberg

1 Introduction
Over the past two decades cognitive scientists have become increasingly interested in how the study of children and adults with neurodevelopmental disorders might inform theories of neurocognitive architecture. This interest has not been without controversy and the extent to which research on neurodevelopmental disorders provides new insights for cognitive science has been hotly debated both on theoretical and empirical grounds. In this chapter I provide a brief overview of this controversy and weigh the arguments for and against what we might learn from studying individuals who have fundamental biological impairments. I then discuss the example of research on theory of mind in two different disorders, autism and Williams syndrome (WMS), which has highlighted a number of important aspects of how this core cognitive capacity develops in both normal and atypical populations.

2 Neurodevelopmental Disorders
2.1 Defining Neurodevelopmental Disorders
The majority of neurodevelopmental disorders are caused by genetic abnormalities that may be classified into several categories. These include disorders that result from mutations in a single gene (e.g., phenylketonuria or fragile-X syndrome), and chromosomal disorders in which an entire chromosome (e.g., Down syndrome or Turner syndrome) or segments of a chromosome (e.g., WMS or Prader-Willi syndrome) are either missing or duplicated. A third group of disorders is referred to as polygenic or complex because they are assumed to be caused by several interacting genes. These disorders (e.g., autism, specific language impairment, or dyslexia) typically involve inherited quantitative cognitive, behavioral or personality traits (Tager-Flusberg, 1999a).

Across all neurodevelopmental disorders genetic abnormalities disrupt the normal course of brain development beginning early during the prenatal period. These developmental brain abnormalities lead to distinct cognitive and behavioral phenotypic outcomes, including mental retardation or learning disabilities, which are characteristic of the majority of individuals with specific disorders. In some disorders we find quite unusual profiles of
cognitive functioning, which may include striking differences between specific cognitive domains (Tager-Flusberg, 2003). For example, people with WMS have relatively strong language skills coupled with severely impaired visual-spatial skills whereas in Down syndrome verbal short term memory is significantly impaired in contrast to their spatial skills, which are commensurate to their mental age levels. These contrasting patterns of cognitive skills are sometimes interpreted as evidence for dissociations between different mental processes and brain systems.

2.2 Neurodevelopmental Disorders and Cognitive Science

Within cognitive science there has been a rich and lengthy tradition of using evidence from people with brain damage to inform cognitive theory, including how cognitive systems are structured and organized. The use of adults with acquired lesions in discrete brain regions, which result in disorders such as amnesia, aphasia or agnosia, has enriched our understanding of the memory, language and visual systems (cf. Gazzaniga, 2000). Adult patients with acquired disorders are assumed to have developed normally and reached a mature end state before the lesion damaged their previously intact cognitive system. Much of what we have learned from studies of these adult patients has involved the comparison of patients, often at the individual subject level, who have contrasting lesion sites and demonstrate dissociated deficits (e.g., Dunn & Kirsner, 2003; Gabrieli, Fleischman, Keane, Reminger, & Morell, 1995; Sternberg, 2001).

The assumptions that underlie studies of adult patients with acquired lesions do not hold for people with neurodevelopmental disorders for whom disruptions in early brain development influence not only the end state but also the development of cognitive systems. Children with such disorders provide a window onto the developmental processes that underlie these atypical patterns, but there are no discrete neural lesions comparable to those found in acquired disorders that are associated with particular disorders. Instead, studies using in vivo brain imaging methods with children or adults suggest that there are subtle differences in the volume and morphology of particular structures in both cortical and subcortical regions associated with different disorders (Lyon & Rumsey, 1996; Thatcher, Lyon, Rumsey, & Krasnegor, 1996). The differences between acquired and developmental disorders are as fundamental as the differences between static and dynamic neurological and neurocognitive systems (Oliver, Johnson, Karmiloff-Smith & Pennington, 2000), leading some to question whether evidence from children or adults with neurodevelopmental
disorders can be used to make claims about the architecture of human cognitive systems, especially claims regarding modularity.

Karmiloff-Smith (1997) has articulated most clearly the key differences between adult and developmental disorders. Her main arguments focus on the fact that brain development in neurodevelopmental disorders differs from normal brain development beginning early in neurogenesis (cf. Courchesne, Townsend, & Chase, 1995). Karmiloff-Smith points out that there is no one-to-one mapping between specific genes and specific cognitive abilities. Genetic influences operate on the trajectory of neural development in complex ways that we are only beginning to understand. These are important points that are often missed by cognitive scientists outside the field of neurodevelopmental disorders.

Karmiloff-Smith (1998) further argues that the developing brain is significantly more plastic than we have appreciated from earlier research on adult acquired disorders. On her view, brain specialization or modularity is not the starting point that directs the course of cognitive development; rather it is the product of development (Karmiloff-Smith, Brown, Grice, & Paterson, 2003). In neurodevelopmental disorders, because development has gone awry, cognitive structure and function will also not be normal. There may be cases where behavioral patterns will appear normal, but Karmiloff-Smith argues that even superficially preserved abilities are the result of atypical underlying developmental processes and brain organization (Karmiloff-Smith et al., 2003; Paterson, Brown, Gsodl, Johnson, & Karmiloff-Smith, 1999). These arguments suggest that individuals with neurodevelopmental disorders cannot provide evidence about how cognitive systems are organized but these claims rest on assumptions about the degree and type of plasticity that is available for the development of neurocognitive systems in different populations.

One might question how far we can take the plasticity claim, or put another way, are there more constraints on the organization and structure of the developing brain than Karmiloff-Smith and her colleagues suggest? At some level, this remains an empirical question. Elsewhere I have argued that there are far fewer differences in developmental disorders than others have assumed (Tager-Flusberg, 2000a). For example, as noted earlier, despite some variation in relative size (either smaller or larger) and other surface features, in fact, across a wide range of disorders it is actually quite remarkable how similar the brains of different populations are to one another and to normally developing children. Sensory and motor systems are located in similar cortical brain regions and higher-level cognitive functions are also organized in quite parallel neural structures. While there is some functional variation, this does not go much beyond the degree that is observed in normal people. This
suggests that the brain is a dynamic system that develops along moderately flexible but fairly bounded and directed pathways that are essentially similar in both normal and disordered populations. On this view, neurodevelopmental disorders can provide cognitive science with converging evidence about cognitive organization. Because there are sometimes striking developmental asynchronies (but not aberrant pathways) and protracted developmental periods in children with different disorders, they may reveal aspects of neurocognitive architecture that are not clearly evident in normally developing children in whom there is close synchrony and rapid timing across many developmental domains.

Despite the fact that neurodevelopmental disorders do not create discrete brain lesions it may still be the case that certain cognitive systems or subsystems will be differentially affected by particular genetic abnormalities. This remains an empirical question. It may hold for some disorders but not others, for some cognitive systems but not others. Nevertheless, we should keep Karmiloff-Smith's cautions in mind; especially since the identification of specific cognitive modules or domains that have been disrupted in any specific neurodevelopmental disorder has been more elusive that we had anticipated (e.g., Frith & Happé, 1998).

2.3 Methodological Issues

Studies of neurodevelopmental disorders have been used as evidence for strong theoretical claims even when they have employed highly questionable methodology that would not be viewed as acceptable in other areas of cognitive science. Supporters of widely different theoretical perspectives have been guilty of taking this relatively uncritical view of the literature on neurodevelopmental disorders, including both nativists (e.g., Pinker, 1994) and non-nativists (e.g., Karmiloff-Smith et al., 2003).

One of the major issues that must be addressed in the study of all neurodevelopmental disorders is the significant heterogeneity in phenotype expression within each syndrome, which may be related to non-genetic as well as genetic variation. Significant individual differences in performance are related to variations in the experiences of people with neurodevelopmental disorders. One important experiential factor is the effect of remediation that is likely to influence a person’s performance on cognitive as well as functional neuroimaging tasks. For example, the neural and behavioral outcomes in adults with a history of severe dyslexia differ considerably for those who have been able to compensate as a result of extensive intervention compared to those who remain severely impaired in their reading ability (Shaywitz et al., 2003). Although this is not a factor that can be effectively controlled
in all studies, it needs to be considered as a potential contributor to performance variation within any study sample. A second factor is age: older people have had significantly more cognitive experience than younger people, which contributes to their task performance. All developmental research acknowledges this important influence, yet it is not unusual to find studies on neurodevelopmental disorders that include young children as well as adults as subjects in the same experiment (e.g., Deruelle, Mancini, Livet, Cassé-Perrot, & de Schonen, 1999).

The two variables that have the most significant impact on cognitive performance in studies of neurodevelopmental disorders are level of intellectual functioning and language. As noted earlier, most neurodevelopmental disorders include mental retardation as part of the phenotype. However, within every disorder, IQ shows the same bell-shaped distribution of scores as in the normal population, though the average will be significantly lower than 100 and the range will vary (Tager-Flusberg, 1999b). For example, in WMS, the average IQ score is about 60, with some people scoring within the normal range of 85 – 100, while other have scores 40 or below. The range is even wider in autism spectrum disorders: about 70% of the autism population have mental retardation (Bailey Phillips & Rutter, 1996), but the full range may be from below 20 to above 150. IQ significantly impacts performance across all cognitive tasks. Its influence is on general rather than domain-specific cognitive processes (e.g., attention, memory, processing speed) and may be linked to cognitive or neural efficiency (Detterman, 1999). Language skills also significantly impact task performance. Many people with neurodevelopmental disorders are language impaired, performing below expectations for age on standard tests of linguistic ability. Impaired language may affect comprehension of task instructions, the content of verbal tasks, or the ability to use language directly or indirectly in responding to an experimental task.

The cumulative effects of IQ and language variation can potentially obscure syndrome-specific patterns of performance across cognitive tasks. In turn, this will have considerable impact on the inferences that can be drawn from studies of neurodevelopmental disorders regarding cognitive structure and organization. Specifically, these non-modular influences on task performance may over-ride what might be revealed about the modular architecture of cognitive systems that have been differentially affected by neurogenetic factors.

The methodological challenges posed by investigating neurodevelopmental disorders cannot be disregarded or completely controlled. We are limited by the availability of study participants drawing from syndromes that are often quite rare. Nevertheless, well-controlled
studies will select experimental subjects representing a relatively narrow age range, document the IQ and language skills of their participants, and include control tasks or comparison groups that are appropriately matched on these key variables. (cf. Tager-Flusberg, Plesa-Skwerer, Faja, & Joseph, 2003).

3 Theory of Mind in Neurodevelopmental Disorders

3.1 Background on Theory of Mind
Successful social interactions depend on the ability to understand other people’s behavior in terms of their mental states, such as beliefs, desires, knowledge, and intentions. Social situations and events cannot be interpreted on the basis of overt behavior without representing the mental states underlying people’s actions. Understanding people as intentional, mental beings is at the core of social cognition, within which the ability to interpret people’s behavior in a mentalistic explanatory framework using a coherent, causally related set of mental constructs is central to a theory of mind (Astington, Harris, & Olson, 1988; Carruthers & Smith, 1996; Perner, 1991; Wellman, 1990; Whiten, 1991).

The past two decades witnessed an exponential increase in research on the development of theory of mind. Studies suggest that the earliest signs of social understanding appear in infancy, including the ability to detect biological motion, goals and intentions, preferential gaze toward people rather than objects, imitation, and joint attention (e.g., Baldwin & Moses, 1996; Csibra, 2003; Hood, Willen, & Driver, 1998; Meltzoff, 1995; Meltzoff & Decety, 2003; Repacholi, 1998; Woodward, 1998; see Chapter by Johnson, this volume). By the time children reach the preschool years they understand mental concepts and are able to predict and explain human actions by inferring the contents of people’s mental states. Understanding that a person’s behavior can be interpreted on the basis of the person’s belief about a situation, which may differ from reality (i.e., a false belief), has been considered a hallmark of a representational theory of mind, which is based on the capacity for metarepresentation. Other types of evidence for a mentalistic construal of persons emerging in the preschool years consist of children’s capacity to use mental states (e.g., desire and belief terms) to explain human action (Bartsch & Wellman, 1995), to use information about a person’s perceptual access or knowledge to judge whether an action was intended or accidental (Schult & Wellman, 1997; Pillow, 1988), and children’s preference for psychological explanations over behavioral descriptions of action scenarios (Lillard & Flavell, 1990; Tager-Flusberg & Sullivan, 1994). These abilities are the main ingredients of a
theory of mind, and have been shown to develop in normally developing children, between
two and five years of age.

Since entering the “cannon” of cognitive systems, findings from studies of
neurodevelopmental disorders have been taken as central in arguments regarding the
structure, architecture, and neural basis of theory of mind (see Baron-Cohen, Tager-Flusberg,
& Cohen, 2000). This began with the influential studies of theory of mind in autism (Baron-
Cohen, Leslie, & Frith, 1985) and has continued more recently with investigations of theory
of mind in WMS (Karmiloff-Smith, Klima, Bellugi, Grant, & Baron-Cohen, 1995; Tager-
Flusberg & Sullivan, 2000), specific language impairment (Miller, 2001), and even deaf (de
Villiers & de Villiers, 2000; Peterson & Siegel, 1995; 1998) and blind (Brown, Hobson, Lee,
& Stevenson, 1997) children. The findings from these studies have been the subject of
considerable controversy, in part because of the methodological problems discussed earlier.
In the remainder of this chapter I review evidence from studies of autism and WMS in favor
of a new model for the organization of theory of mind, with special attention to how
particular components of this cognitive system may be integrally linked to specific aspects of
language.

3.1 Theory of Mind in Autism

Baron-Cohen and his colleagues were the first to demonstrate that the majority of children
with autism failed false belief tasks, in contrast to normally developing preschoolers and
children with Down syndrome (Baron-Cohen et al., 1985). Follow-up experimental studies
provided further support for their hypothesis that autistic children are impaired in their
acquisition of a theory of mind: they fail to understand stories that involve deception or false
belief (Baron-Cohen, Leslie, & Frith, 1986), they do not understand the connection between
perception and knowledge (Baron-Cohen 1989), they lack imagination (Scott & Baron-
Cohen, 1996) and do not engage in spontaneous pretend play (Baron-Cohen, 1987; Lewis &
Boucher, 1988). In naturalistic settings children with autism do not use mental state terms
such as think and know in everyday conversation (Tager-Flusberg, 1992), and lack social
skills that depend on mentalizing (Frith, Happé, and Siddons, 1994; for review see Baron-
Cohen, 2000).

The significance of the theory of mind hypothesis of autism, as it came to be known
in the literature (Baron-Cohen, Tager-Flusberg, & Cohen, 1993), was that it not only
explained the failure of children with autism on tasks tapping theory of mind abilities, but
also provided a unified cognitive explanation for the primary diagnostic impairments in
pretence, social functioning, and communication (Baron-Cohen, 1988; Frith, 1989; Leslie, 1987). It revolutionized research on autism, and had important influences on theoretical models of theory of mind. The selective deficits in theory of mind among people with autism was taken as evidence in support of the modularity of theory of mind (e.g., Baron-Cohen, 1995; Leslie & Roth, 1993). More recent neuroimaging studies also suggest that the neural circuits that subserve theory of mind may be fundamentally impaired in autism (e.g., Frith & Frith, 1999; 2000).

3.2 Theory of Mind in Williams Syndrome

WMS is a rare neurodevelopmental disorder caused by a hemizygous microdeletion on the long arm of chromosome 7 (7q11.32), which includes between 16 and 25 genes (Bellugi, Lichtenberger, Mills, Galaburda & Korenberg, 1999; Osborne & Pober, 2001). The syndrome is characterized by a unique phenotype that typically includes physiological abnormalities of the heart and other organs, a variety of connective or soft tissue disorders, cranio-facial dysmorphology and an unusual combination of cognitive, personality and behavioral features (Morris & Mervis, 1999).

Although the majority of individuals with WMS are mentally retarded, some aspects of their cognitive functioning appear relatively spared, including vocabulary knowledge (Bellugi, Bihle, Neville, & Doherty, 1992; Mervis et al., 1999; Volterra, Capirci, Pezzini, Sabbadini, & Vicari, 1996), face processing (Bellugi, Marks, Bihrle & Sabo, 1988; Tager-Flusberg et al., 2003), and auditory rote memory (Mervis et al., 1999; Robinson, Mervis & Robinson, 2003 Udwin & Yule, 1991). The behavior and personality of people with WMS also suggest some unique characteristics. The most remarkable feature of both children and adults with WMS is their extreme interest in people. They have a warm, outgoing, cheerful and friendly personality style (Klein-Tasman & Mervis, 2003; Udwin & Yule, 1991). They are described as being empathic towards other people (Gosch & Pankau, 1994), they are less reserved toward strangers, more approaching, curious and extroverted, and overly friendly and affectionate (Gosch & Pankau, 1997; Sarimski, 1997; Tomc, Williamson, & Pauli, 1990; Van Lieshout, De Meyer, Curfs and Fryns, 1998).

The cognitive and personality profile associated with WMS led several researchers to propose that WMS may be characterized by sparing in the domain of theory of mind (Karmiloff-Smith, et al., 1995; Tager-Flusberg, Boshart, & Baron-Cohen, 1998). The combination of relatively good language skills, excellent face processing abilities, strong social interest, and attention to faces and people (Mervis et al., 2003), helped to foster the
view that theory of mind might be spared in this population. The initial evidence came from a study by Karmiloff-Smith and her colleagues (1995) who used a set of standard theory of mind tests including first- and second-order false belief tasks and a higher-order task that involved attributing intentions to linguistic utterances. Karmiloff-Smith and her colleagues found that the majority of the subjects with WMS passed the first-order tasks, and some even passed the second- and higher-order tasks. They concluded from their findings that WMS involves an “islet of relatively preserved ability” (Karmiloff-Smith et al., 1995, p. 202) in theory of mind.

Of particular interest to theory of mind scholars is the striking contrast between autism and WMS. In autism there are fundamental impairments in language (Kjelgaard & Tager-Flusberg, 2001; Lord & Paul, 1997), face processing (e.g., Langdell, 1978; Joseph & Tanaka, 2003), and severe social deficits (Klin, Schultz, & Cohen, 2000); exactly those cognitive skills that are relatively preserved in WMS. Furthermore, in autism visual-spatial skills (as measured for example by block design tasks) are spared (e.g., Joseph, Tager-Flusberg & Lord, 2002; Shah & Frith, 1993) but are severely impaired in WMS (Bellugi, Sabo, & Vaid, 1988; Hoffman, Landau, & Pagani, 2003; Mervis, Robinson, Bertrand, Morris, Klein-Tasman, & Armstrong, 2000). These contrasting profiles suggest a double dissociation: in autism theory of mind is impaired while visual-spatial skills are spared; in WMS theory of mind is spared while visual-spatial skills are impaired. These arguments provide support for the view that these domains are separable in terms of their underlying cognitive and neural mechanisms.

3.3 Criticisms of the Theory of Mind Hypothesis in Autism and WMS

Despite its wide-ranging appeal, the theory of mind hypothesis of autism has come under attack (see Tager-Flusberg, 2001). Researchers have questioned the selectivity or uniqueness of theory of mind impairments in autism, because studies show that non-autistic children and adolescents with mental retardation also fail standard theory of mind tasks at a higher rate than would be expected give their age and developmental level (Benson, Abbeduto, Short, Bibler-Nuccio, & Maas, 1993; Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998; Zelazo, Burack, Benedetto, & Frye, 1996). The same is true for other populations such as oral deaf children (de Villiers & de Villiers, 2000; Peterson & Siegel, 1995; 1998), and people with schizophrenia (Corcoran, 2000). If these groups also have difficulty on theory of mind tasks, can theory of mind be interpreted as the unique deficit in autism?
Other concerns voiced by some researchers include the fact that autism symptoms emerge in infancy long before normally developing children would pass theory of mind tasks, such as false belief (Klin & Volkmar, 1993). Furthermore, there are features of autism that are not so clearly interpreted in terms of a core cognitive impairment in theory of mind. These include the primary symptoms of repetitive behavior and restricted or obsessive interests (Turner, 1999) and other secondary features such as savant abilities (such as outstanding memory for facts, perfect pitch, calendrical calculators, or artistic talent), deficits in the ability to generalize, exceptionally good visual perceptual skills, and atypical sensory sensitivities. Impairments in theory of mind do not explain these features of the disorder (Happé, 1999; Plaisted, 2000).

One final criticism of the theory of mind hypothesis comes from the fact that some children with autism pass theory of mind tasks, including false belief (Baron-Cohen et al., 1985). The numbers who pass varies from one study to the next, but even a small percentage (e.g., 20% in Baron-Cohen et al., 1995) must be accounted for in any theory. If autism involves a failure to develop a theory of mind, the question is, how could these participants with autism pass the tasks? One explanation is that theory of mind may be seriously delayed in autism, and most people never achieve the same endpoint as non-autistic people. Others argue that in autism failure on tasks that tap theory of mind abilities may be more directly interpreted in terms of domain-general deficits in either executive functions (e.g., Russell, 1997) or language (e.g., Eisenmajer & Prior, 1991; Tager-Flusberg, 2000b). This latter argument challenge the view that difficulties on false belief and related tasks directly reflect domain-specific impairments to theory of mind.

At the same time, recent evidence suggests that theory of mind may not be as spared in WMS as originally believed. We systematically investigated performance on false belief and other theory of mind tasks in children with WMS. The children with WMS were matched to two comparison groups on age (4 to 10 years), IQ, and standardized language measures. The comparison groups included children with Prader-Willi syndrome (PWS), another genetically based neurodevelopmental disorder, and children with non-specific mental retardation. In each experiment between 15 and 25 children were included in each group. On three different first-order theory of mind tasks, false belief, explanation of action (Tager-Flusberg & Sullivan, 1994), and understanding of intended action (Joseph & Tager-Flusberg, 1999), we found that the children with WMS performed no better than the matched comparison groups (Plesa-Skwerer & Tager-Flusberg, in press; Tager-Flusberg & Sullivan, 2000).
We also investigated higher-order theory of mind tasks in adolescents with WMS, and matched groups of adolescents with PWS and mental retardation. Again no differences were found among these groups in second-order belief reasoning (Sullivan & Tager-Flusberg, 1999), distinguishing lies and jokes (Sullivan, Winner, & Tager-Flusberg, 2003), or in using trait information to attribute intentionality (Plesa-Skwerer & Tager-Flusberg, in press). Thus our more recent studies on theory of mind in WMS provide no evidence of relative sparing in this domain for either children or adolescents with WMS compared to age, IQ, and language-matched controls (cf. earlier discussion on Methodology Issues, Section 2.3).

4 Model of Theory of Mind

4.1 Two-Component Model

Research on theory of mind in autism and Williams syndrome has left us with contradictory hypotheses and findings. On the one hand, it has been proposed that these disorders represent dissociation in theory of mind abilities, reflecting their contrasting social profiles. On the other hand, current data suggest that there may be little to distinguish between these groups in their performance on theory of mind tasks (see also Pearlman-Avnion, 2003). To resolve this apparent paradox we have recently proposed a componential model of theory of mind in which there are several interacting hierarchically organized component levels within the domain of theory of mind (Tager-Flusberg, 2001; Tager-Flusberg & Sullivan, 2000). The advantages of this model are that it accounts for the broader range of phenomena that are now encompassed by conceptions of theory of mind or mentalizing than original theoretical accounts (e.g., Leslie & Thaiss, 1992): it incorporates known developmental aspects of theory of mind from infancy through middle childhood; it is consistent with neurobiological models of social neuroscience; and it can explain the pattern of findings from autism and WMS (as well as other disordered populations). The model presented here is quite preliminary; I limit discussion to two key components of theory of mind: the perceptual, and cognitive components. No doubt as research in this area advances these will be further divided into additional discrete components.

On this model there are two levels or components where mental states are represented: a primary social-perceptual level and a higher-order social-cognitive level. The perceptual component refers to the on-line immediate or intuitive representation of a person’s mental state, based on information directly available in faces, voices, and body posture and movement. The cognitive component refers to our metarepresentational capacity to make more complex cognitive inferences about the content of mental states that requires integrating
information across time and events. This distinction between perceptual and cognitive levels of representation corresponds roughly to the categories of “intuitive belief” and “reflective beliefs” proposed by Sperber (1997)\(^1\).

### 4.2 Social-Perceptual Component

The social-perceptual component of theory of mind builds on the innate preferences of infants to attend to human social stimuli, especially faces and voices (e.g., Fernald, 1989, 1993; Johnson & Morton, 1991; Mehler & Dupoux, 1994). The route to interpreting mental state information from these stimuli lies in the interaction of innately specified mechanisms with social information in the world, which is obtained through continued interactions with people. The social preferences of infants that promote continued interactions with people might be driven by affective motives – the intrinsic reward of social stimuli. By the latter half of the first year of life infants use perceptual information from faces, voices and gestures to interpret the intentions and emotional states of other people; they may also use more subtle cues such as eye gaze to judge what another person is attending to or planning to do (cf. Baldwin, 1993; Baron-Cohen, 1994; Repacholi, 1998). Thus, the perceptual component of theory of mind emerges first in development, and is available to infants for making a range of mental state judgments about other people based primarily on sensory inputs. Over the course of development social perceptual judgements may also entail other cognitive inputs (e.g., memory).

### 4.3 Social-Cognitive Component

The social-cognitive component of theory of mind builds on the earlier emerging perceptual component. This component is involved in making mental state inferences that depend on integrating information not only from perceptual cues, but also from sequences of events over time. The social-cognitive component of theory of mind is more closely linked to other cognitive or information processing systems, such as working memory (needed for integrating information) and language. The development of the cognitive component of theory of mind begins during the early preschool years when children begin to talk and reason about epistemic states (Bartsch & Wellman, 1995). It is firmly in place by four years.

\(^1\) There have been other researchers who have proposed more complex models of theory of mind, for example, Wellman (1990) and Baron-Cohen (1995). Their models share certain features with the model presented here, but they also differ in terms of either the nature of the core components (Wellman) or the more limited developmental framework (Baron-Cohen, 1995) that is endorsed here. The model presented here was explicitly developed to account for the detailed evidence of both spared and impaired aspects of mentalizing found in autism and Williams syndrome (Tager-Flusberg, 2001; Sullivan & Tager-Flusberg, 2000).
of age, when young children have the metarepresentational capacity to pass false belief and other related tasks. Language plays an especially significant role in the development of this component of theory of mind (de Villiers, 2000; Hale & Tager-Flusberg, 2003).

4.4 Neurobiological Evidence

In this model, the two main components to a theory of mind, each has its own developmental time course, and each is dependent on different underlying neurocognitive mechanisms. Converging evidence comes from studies of brain function, particularly from research on the neurobiological substrate of what Leslie Brothers (1990) refers to as the “social brain.” The primary areas of the brain that are involved in social-perceptual information processing include the amygdala and associated regions of medial temporal cortex, including the superior temporal sulcus (Allison, Puce, & McGarthy, 2000). The amygdala is central to the processing of emotion (e.g., Adolphs, Tranel, Damasio, & Damasio, 1994; Adolphs, Tranel & Damasio, 1998) and other complex social stimuli (Brothers, Ring, & Kling, 1990; Perrett et al., 1990). Functional brain imaging studies show that the amygdala and areas of the medial temporal cortex are activated in tasks tapping the recognition of facial expressions of emotions and other mental states (Baron-Cohen et al., 1999; Breiter et al., 1996) as well as the perception of biological or intentional motion (Bonda, Petrides, Ostry, & Evans, 1996).

The brain areas that subserve the social-cognitive component of theory of mind include regions in the prefrontal cortex. The orbito-frontal cortex is involved in reasoning about the social appropriateness of action (Eslinger & Damasio, 1985) and in making lexical judgements about cognitive mental state terms (Baron-Cohen, Ring, Moriarty, Schmitz, Costa, & Ell, 1994). Areas in the medial frontal cortex are closely associated with other theory of mind abilities, especially tasks tapping advanced social-cognitive capacities (Fletcher et al., 1995; Goel, Grafman, Sadato, & Hallett, 1995). In summary, there is preliminary evidence that different neural substrates underlie the components of theory of mind described here (see also Frith & Frith, 2003; Siegal & Varley, 2002). These brain regions form a complex neural circuit for processing a range of social information from basic perception of biological motion to inferring the contents of other people’s minds.

5 Application of the Componential Model to Neurodevelopmental Disorders

5.1 Theory of Mind Deficits in Autism

The componential model provides for two levels of representing mental states: the perceptual and cognitive. The perceptual component is the primary level in that it directly computes
mental states on the basis of available information, it emerges early in development and is based in both subcortical as well as cortical brain regions. The cognitive component, while it builds on the perceptual level and is closely interconnected, is a higher-order level, interacting with other cognitive systems including memory and language to compute the contents of mental states in prefrontal cortical regions. On this view, theory of mind is conceived in broader terms that the original metarepresentational theories (cf. Leslie & Roth, 1993). Within this model the fundamental domain-specific deficits in autism are in the social-perceptual component of theory of mind: children and adults with autism are fundamentally impaired in computing mental states on the basis of information available from social stimuli, especially faces and voices.

The roots of the social-perceptual impairments in autism may be seen in the social orienting deficits that are evident in infants (Dawson, Meltzoff, Osterling, Rinaldi & Brown, 1998; Klin, 1991; Osterling & Dawson, 1994). These deficits are correlated with their failure to perceive behavior in others as intentional or to appreciate others’ perspectives, as exemplified in the joint attention deficits that are among the hallmark symptoms of the disorder (Mundy & Sigman, 1989; Mundy, Sigman, & Kasari, 1990, 1993). Thus, children with autism below the age of three demonstrate significant impairment in the range of behaviors that are among the early developments in the social-perceptual component of theory of mind (cf. Klin & Volkmar, 1993). Even older high-functioning people with autism or Asperger syndrome perform poorly on tasks that measure the perception of biological motion (Blake, Turner, Smoski, Pozdol, & Stone, 2003), the ability to read mental states from the eye region of the face (Baron-Cohen, Joliffe, Mortimore, & Robinson, 1997) or the attribution of intentional and social significance to ambiguous visual stimuli (Klin, Schultz & Cohen, 2000; Klin, Jones, Schultz & Volkmar, 2003).

The majority of children with autism are also impaired on social-cognitive measures of theory of mind, as evidenced by their failure to pass false belief tasks (Baron-Cohen et al., 1985), or to explain human behavior using mental state terms (Tager-Flusberg, 1992). These deficits in the cognitive aspects of theory of mind grow out of the earlier deficits in social-perception because these components are closely interconnected, with cognition building on social perception. Nevertheless, as noted earlier, some children with autism pass theory of mind tasks. Within the componential framework, I argue that these children depend on language (not theory of mind) to hack out solutions to such tasks, which they treat as logical problems (Tager-Flusberg, 2001). At the same time these children remain fundamentally impaired at the social-perceptual level, and in their conceptual understanding of mental states.
In a later section I provide evidence for this claim that children with autism can pass theory of mind tasks via language.

5.2 Theory of Mind Deficits in Williams Syndrome

Children and adults with WMS are fundamentally different from people with autism in that they show an extremely strong interest in and sensitivity to others (Jones et al., 2000). Based on evidence from two preliminary studies we argued that these aspects of their behavior reflect relative sparing in the social-perceptual component of theory of mind (Tager-Flusberg & Sullivan, 2000). On this view, the dissociation between autism and WMS is at the social-perceptual level of representing mental states. At the same, there is no dissociation between autism and WMS at the social-cognitive level; both groups generally perform poorly on classic theory of mind tasks relative to age-matched peers. Thus, there is no double dissociation between these populations within these two components of theory of mind.

We conducted two small-scale studies on the social-perceptual component of theory of mind in WMS. Tager-Flusberg et al. (1998) compared adults with WMS to a well-matched group of adults with Prader-Willi syndrome on the Eyes task (Baron-Cohen et al., 1997), for which a subject is asked to select which of two terms best describes the mental state expressed in a photograph of the eye region of a face. At noted earlier, Baron-Cohen et al. (1997) found that high-functioning adults with autism performed significantly worse than controls on this task. In our study the adults with WMS performed significantly better than the adults with Prader-Willi syndrome. In fact half the WMS group performed at the same level as normal age-matched adults. These findings were taken as evidence that WMS involves sparing in theory of mind; some people with WMS may be spared in the absolute sense (i.e., those performing within the limits of the normal population) while other were spared in the relative sense (compared to matched adults with Prader-Willi syndrome). In another study Tager-Flusberg and Sullivan (1999) found that young children with WMS showed significantly greater empathy than a matched group of children with Prader-Willi syndrome. Their task involved comparing the verbal and non-verbal responses of the subjects to the distress exhibited by an experimenter when she feigned hurting her knee. The children with WMS showed greater concern, more appropriate affect, and made more relevant verbal empathic comments than the comparison group. Both of these studies involved measures of the social-perceptual component of theory of mind in that they tap the ability to read facial expressions of mental states rather than the ability to make inferences about the contents of another person’s mind.
We are currently following up on these earlier studies using better-controlled tasks and larger, more heterogeneous, groups of adolescents and adults with WMS. Based on preliminary analyses of our data it is no longer so clear that people with WMS are spared in their ability to compute mental state information from facial or vocal expressions. Across several experiments the WMS subjects perform worse than normal controls, although their pattern of performance is similar, suggesting the use of the same cognitive mechanisms (Plesa-Skwerer, Faja, & Tager-Flusberg, 2003). Thus, it seems that even social-perceptual theory of mind tasks entail some domain-general processing skills such as attention or response speed, which are most likely to be compromised in any person with mental retardation. It remains to be seen whether the unusual sociability that is a central feature of the WMS phenotype is related to theory of mind, or is a reflection of unique arousal and emotional functioning.

6 Language and Theory of Mind

6.1 Language and the Social-Cognitive Component

One of the fundamental differences between the perceptual and cognitive components of theory of mind lies in the role of language. The social-cognitive component is integrally linked to language. Evidence for this close relationship between language and a representational understanding of mind comes from several sources, including developmental studies of preschoolers. Numerous studies have found a significant correlation between standardized language measures and performance on theory of mind tasks in preschoolers (e.g., Cutting & Dunn, 1999; Hughes & Dunn, 1997; Jenkins & Astington, 1996). Astington and Jenkins (1999) conducted a longitudinal study in order to identify the direction of this relationship. Their findings confirmed that language predicted later performance on theory of mind tasks but not the reverse. Furthermore, syntactic knowledge was the major factor predicting to later theory of mind.

What is significant about syntax in relation to false belief? de Villiers and her colleagues (de Villiers, 2000; de Villiers & de Villiers, 2000, de Villiers & Pyers, 2002) argue that sentential (or tensed) complements are a prerequisite to the child’s acquisition of a representational theory of mind. Sentential complements, which allow for the embedding of tensed propositions under a main verb, have unique syntactic and semantic properties. Two classes of verbs take sentential complements: verbs of communication (e.g., John said that Fred went shopping) and verbs of mental state (e.g., Mary thought that Fred went to the movies). In sentential complements the embedded clause is an obligatory linguistic argument
that may have an independent truth value. Therefore, the main clause may be true (e.g., John said X; Mary thought Y) while the embedded clause may be false (e.g., Fred did not go either shopping or to the movies). The syntax and semantics of sentential complements allow for the explicit or meta-representation of a falsely embedded proposition.

A few studies have documented a significant correlation between knowledge of sentential complements and performance on theory of mind tasks in preschool-aged children (e.g., de Villiers & Pyers, 2002; Tager-Flusberg, 1997, 2000b). In a longitudinal study, carried out over the course of a year, de Villiers and Pyers (2002) found that mastery of sentential complements predicted later theory of mind performance independent of general language change, but that the reverse did not hold. Two recent training studies have provided further evidence that explicit training on the syntax of sentential complements promotes the acquisition of false belief (Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003), even when mental state verbs were not incorporated into the training phase of the study. These studies highlight the significance of language, specifically sentential complements in acquiring the social-cognitive component of theory of mind. At the same time, contrary to de Villiers’ (2000) predictions, the training studies demonstrated that for normally developing preschoolers acquisition of sentential complements was not a necessary prerequisite to passing false belief tasks. Hale and Tager-Flusberg (2003) included a group of children who were trained only on the false belief task (not including complex complement constructions). After the training phase these children performed as well as the group of children trained on complements on the theory of mind post-tests, however they had not mastered sentential complements. Similarly, Lohmann and Tomasello (2003) found significant advances in theory of mind abilities in a group of children whose training consisted of exposure to perspective-shifting discourse (again, no complex syntax) in the context of deceptive objects. Thus while sentential complements may strongly facilitate the acquisition of theory of mind because the linguistic representations for complements are isomorphic to the representational format needed for propositional attitudes (see de Villiers, 2000), they do not constitute the sole developmental pathway to achieving this new level of metarepresentational capacity.

6.2 Language and Theory of Mind in Neurodevelopmental Disorders

In this section I discuss how the componential model, specifically the role of language in the social-cognitive component, might help to reconcile the contradictory perspectives on theory of mind in children with autism and Williams syndrome. Recall that one major concern about the theory of mind hypothesis of autism is that it does not explain why some children with
Neurodevelopmental Disorders and Cognitive Science 18

autism pass false belief tasks. At the same time, our data showed that contrary to initial speculation, children with Williams syndrome do not perform especially well on such tasks, compared to matched control groups. If, as argued in the previous sections, language plays a significant role in the social-cognitive component, exemplified by false belief tasks, then we should find that variation in theory of mind performance in both autism and WMS is predicted by linguistic ability.

In autism, deficits in pragmatic aspects of language communication are universal, and are among the core diagnostic symptoms that define the syndrome. At the same time, linguistic development is much more variable. About 25% of high-functioning children with autism acquire a rich vocabulary and fully master the grammar of their native language, while about 75% of verbal high-functioning children remain language impaired to different degrees of severity (Kjelgaard & Tager-Flusberg, 2001). If language plays an important determining role in performance on theory of mind tasks, then this variability in linguistic ability may explain why some children with autism are able to pass false belief tasks: they do so via language.

Across many studies on false belief in children with autism, performance is significantly correlated with standardized measures of language, including measures of verbal mental age and pragmatics (Eisenmajer & Prior, 1991), vocabulary (Dahlgren & Trillingsgaard, 1996; Happé, 1995; Sparrevohn & Howie, 1995; Tager-Flusberg & Sullivan, 1994) and syntax (Fisher, 2002; Tager-Flusberg & Sullivan, 1994). We followed up these studies on the influences of general language measures with both cross-sectional and longitudinal studies that investigated whether this relationship between language and theory in autism was more specifically related to mastery of sentential complements. Tager-Flusberg (2000b) compared autistic, and age, IQ and language-matched mentally retarded adolescents on three experiments that tested knowledge of the syntactic and semantic properties sentential complement constructions, including both communication and mental state verbs. In all three experiments performance by participants in both groups on the complementation tasks was significantly related to whether they passed or failed false belief, and in regression analyses, complement knowledge was the single best predictor (over and above IQ and general language measures) of performance on the false belief task. However, for the autistic subjects, false belief was only significantly related to performance on communication verbs, and, in contrast to the mentally retarded group, they showed little sensitivity to the conceptual or linguistic properties of the cognitive mental state verbs used in these studies.
In a more recent longitudinal study of over 50 children with autism between the ages of 5 and 14, we replicated these findings (Tager-Flusberg & Joseph, in press). At two time points, spaced about one year apart, we collected data on theory of mind performance, comprehension of sentential complements for communication and mental state verbs, as well as measures of general language ability. The cross-sectional data collected at the first time point found that knowledge of complements, but only for communication verbs, accounted for 25% of the variance in theory of mind score, beyond the variance explained by age and general language level (which accounted for 43% of the variance). The longitudinal data were analyzed looking at which variables at the first time point predicted theory of mind one year later. Not surprisingly, general language and the children’s original theory of mind scores accounted for 74% of the variance in later theory of mind. Again, additional unique variance (about 8%) was accounted for by performance on the sentential complements task for communication verbs.

These findings show that children with autism who have more advanced language skills, specifically those who have acquired sentential complements for communication verbs, are able to use this linguistic knowledge to master tasks that tap the social-cognitive component of theory of mind. Some children with autism, the minority with normal or near-normal linguistic ability, can use language to reason logically through false belief tasks. Based on this evidence, we have argued in contrast to normally developing preschoolers, language provides the sole route to understanding propositional attitudes; for people with autism there is no independent language-of-thought in the domain of theory of mind. There is some evidence from functional imaging studies using theory of mind tasks that higher-functioning adults with autism activate brain regions that are not typically associated with theory of mind suggesting that they are using different neurocognitive mechanisms (see Frith & Frith, 2000).

Turning now to children with Williams syndrome: why did they perform relatively poorly on false belief and other related tasks (Tager-Flusberg & Sullivan, 2000)? In our studies we carefully matched the children with WMS to two other groups of mentally retarded children, one with Prader-Willi syndrome and one with unspecified retardation. All three groups were matched on age, IQ and language, including both vocabulary and syntactic measures. Because performance on theory of mind tasks is so closely linked to language, it is not surprising that the matched WMS children did not perform better than the other retarded children. Indeed, these findings provide further evidence for the significant role that language (and perhaps other domain-general cognitive processes) plays in solving classic theory of
mind tasks. In these studies on theory of mind, we gave a small number of children with WMS a sentential complements task (the same one used in Tager-Flusberg, 2000b). Although our sample is not large, within this group of 10 children we found that mastery of sentential complements was highly correlated with performance on the false belief tasks, replicating our earlier cross-sectional work on autism. Taken together, the studies summarized here provide strong evidence from normally developing children, children with autism and children with WMS to support for the claim that language, especially sentential complements, plays an important role in the acquisition of a representational theory of mind.

7 Conclusions

Our studies on theory of mind in autism, WMS, and other comparison groups with neurodevelopmental disorders have led us to new insights about the cognitive structure and organization of this important domain of human cognition. Nevertheless, it is important to stress the fact that the componential model outlined here is based on converging evidence not only from the study of autism and Williams syndrome, but also from developmental and neurobiological research. Our claim for the distinction between the perceptual and cognitive levels for representing mental states is also consistent with other accounts of the hierarchical nature of representational systems (Sperber, 1997), including the significant role played by language at the metarepresentational level (Sperber, 2000).

The research program presented here supports the view that we can indeed learn from well-designed studies involving children with neurodevelopmental disorders, though we may easily be led astray if our research designs are not adequate. Despite the fact that children with autism or WMS have genetic abnormalities that no doubt influence brain development in significant ways, there are still important and consistent similarities found across all children. At the same time, unusual phenotypic profiles reveal a great deal about the mechanisms that underlie cognitive performance. Thus, our studies have led us to hypothesize that autism involves selective impairments to those brain areas that are critical for perceiving and representing socially relevant information such as biological motion, facial and vocal expression. In WMS there is no such impairment, though these brain areas may also not be specifically spared – the evidence thus far is equivocal. Future research with these and other populations will ultimately provide a more complete and detailed picture of the neurocognitive architecture of theory of mind; our componential model is simply a rudimentary sketch.
Acknowledgements

Preparation of this chapter was supported by grants from the National Institutes of Health (PO1/U19 DC 03610 and RO1 HD 33470).

References


Language development in exceptional circumstances (pp. 177-189). Edinburgh: Churchill Livingstone.


