

Evaluating the Theory-of-Mind Hypothesis of Autism

Helen Tager-Flusberg

Boston University School of Medicine

ABSTRACT—*Two decades ago, the theory-of-mind hypothesis of autism was introduced by Baron-Cohen and his colleagues; this theory provided a unified cognitive explanation for the key social and communication symptoms in that disorder. I evaluate the theory-of-mind hypothesis in light of studies that have addressed several key questions: Do children with autism develop theory-of-mind concepts? How can we explain why some children with autism pass theory-of-mind tasks? Do deficits in theory of mind account for the major impairments that characterize autistic disorder? Current research supports the view that autism involves delays and deficits not only in the development of a theory of mind but also in additional aspects of social-affective information processing that extend beyond the traditional boundaries of theory of mind.*

KEYWORDS—*autism; theory of mind; social cognition; communication*

Daily social life depends on the ability to evaluate the behavior of other people on the basis of their mental states, such as their goals, emotions, and beliefs. This is accomplished by dedicated cognitive systems, collectively referred to as *theory of mind*. By age four, children normally pass tasks that tap mental-state understanding, including the hallmark false-belief tasks that require a child to distinguish between the world as it really is and the way it might be represented (incorrectly) in the mind of another person. In the classic Sally–Anne false-belief task, a child is told the following story, accompanied by supporting pictures or toy props: Sally places her ball in a basket and goes out to play; while she is gone, Anne takes the ball from the basket and hides it inside a box. The child is then asked where Sally will look for the hidden ball (or where she thinks it will be located) when she returns to play with it again (see Fig. 1).

Address correspondence to Helen Tager-Flusberg, Department of Anatomy & Neurobiology, Boston University School of Medicine, 715 Albany Street L-814, Boston, MA 02118; e-mail: htagerf@bu.edu.

Two decades ago, Baron-Cohen and his colleagues revolutionized autism research when they introduced the theory-of-mind hypothesis to explain the main behavioral symptoms that characterize this neurodevelopmental disorder. Their initial studies showed that most children with autism whose mental and verbal abilities were well beyond the 4-year-old level nevertheless failed the Sally–Anne task and other related tasks (Baron-Cohen, Leslie, & Frith, 1985). Deficits in the acquisition of a theory of mind provided a plausible explanation for the major symptoms of autism, especially impairments in social reciprocity and communication, thus providing the first integrated account of the cognitive mechanisms that might underlie several key behaviors that define the disorder. The original studies have been replicated many times by different research teams, and there is little doubt that children with autism have difficulty attributing mental states to themselves or to other people.

Yet despite the robustness of the empirical findings, there is now less excitement and increasing skepticism among many investigators about the significance of the theory-of-mind hypothesis of autism. Questions about the universality and uniqueness of theory-of-mind impairments in autism, and about how this hypothesis could account for the earliest manifestations of autistic symptoms, have been raised (Tager-Flusberg, 2001). Autism is generally defined on the basis of impairments not only in social and communicative functioning but also in restricted or repetitive behavior patterns. The theory-of-mind hypothesis does not extend to explaining these areas of impairment; nor does it explain some of the strengths that are characteristic of people with autism, such as their superior visual-attention skills.

The past decade has witnessed an exponential increase in research on autism. Today, autism is clearly understood to be a complex and heterogeneous set of related developmental disorders in which no single cognitive mechanism or cause can account for the variety of symptoms and range in their expression. Even the social-communication impairments cannot be explained exclusively on the basis of theory-of-mind impairments. Nevertheless, current research supports the view that children and adults with autism have problems processing mental-state information; that when they are able to infer mental

Summary
*

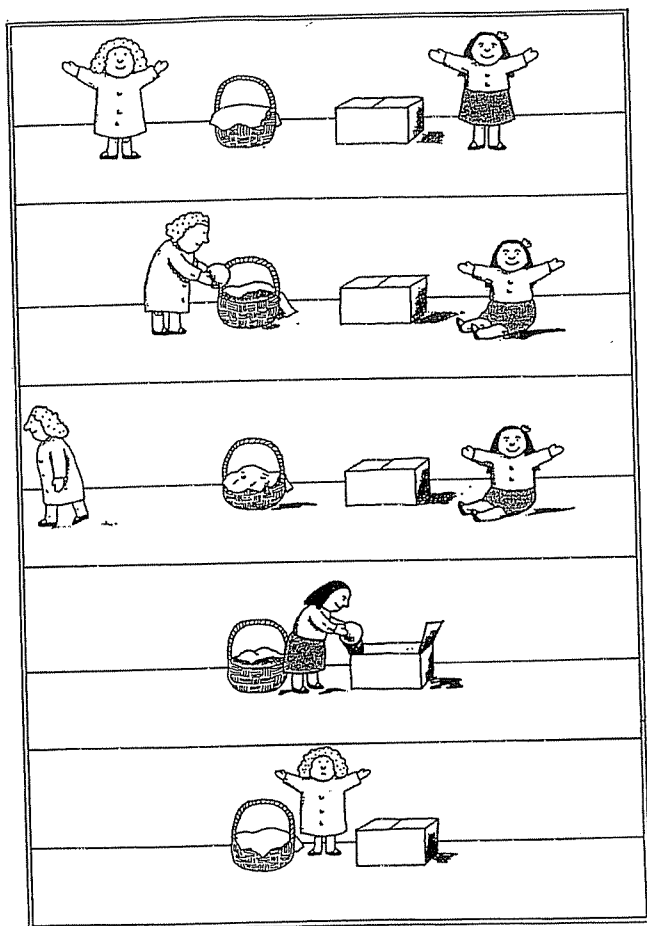


Fig. 1. The Sally-Anne false-belief task. To test theory-of-mind skills, children are presented with a story in which (from the top frame to the bottom) Sally (left) has a basket and Ann (right) has a box; Sally puts her marble in the basket; Sally leaves; Ann takes the marble from the basket and puts it in the box; and Sally returns to look for the marble. Subjects are then asked whether Sally will look for the marble in the basket or the box. Reprinted from *Autism: Explaining the Enigma*, by U. Frith, 1989, Blackwell Publishers, p. 83. Copyright 1989, Blackwell Publishers. Reprinted with permission.

* states, they tend not to use the same neurocognitive systems as do nonautistic people; and that performance on theory-of-mind tasks can account for some, though not all, of the severity of the social and communication symptoms that define this disorder.

A DEVELOPMENTAL PERSPECTIVE ON THEORY OF MIND IN AUTISM

Most studies investigating theory of mind in autism focused on the transition that takes place at the age of four, when children typically first understand false belief. This narrow perspective appeared to reduce a complex social-cognitive developmental progression to a categorical capacity indexed by passing or failing a single task, thus encouraging the notion that autism could be defined as the "absence" of a theory of mind. Yet in all studies that have been conducted, some children with autism

pass false-belief tasks. At the same time, research has shown that older children with different disorders (e.g., nonsigning deaf children; Peterson, Wellman, & Liu, 2005) fail these tasks—evidence that challenges the notion that theory-of-mind deficits are universal and specific to autism.

More recent studies have taken a developmental approach by including a wider range of tasks that tap a child's ability to reason about mental states from the early preschool years to adolescence. For example, Steele, Joseph, and Tager-Flusberg (2003) conducted a longitudinal study in a large sample of children with autism; the participants were administered a battery of nine developmentally sequenced reasoning tasks ranging from tasks appropriate for toddlers (assessing the ability to understand the concept of pretense and the ability to attribute desire to others), young children (assessing the ability to correctly match others' emotions to situations and to understand others' knowledge and false beliefs), or older children and young adolescents (assessing understanding of others' second-order false belief, nonliteral language, traits and intentions, and moral judgments). Over two thirds of the children made some gains over the course of 1 year, and although all the children were delayed relative to their age, a small number of the most able adolescents passed some of the more advanced tasks. Peterson and her colleagues (Peterson et al., 2005) used a set of five tasks that formed a coherent scale of theory-of-mind concepts, from desire to false belief and hidden emotions. The children with autism were generally similar to the other groups in their ability to pass the tasks according to their developmental sequence, but unlike the other groups, children with autism found false belief harder than hidden emotions. While children with autism develop some understanding of desire and emotion, belief and other cognitive states seem to pose a unique challenge to this population.

EXECUTIVE FUNCTIONS, LANGUAGE, AND THEORY OF MIND IN AUTISM

An important question that has generated considerable research focuses on how some children with autism pass theory-of-mind tasks. For nonautistic children, performance on classic theory-of-mind tasks reflects intuitive social insights into people or conceptual knowledge of mental states coupled with general cognitive skills that support the verbal processing, memory of key narrative events, and inhibition of spontaneous responses that are central to the tasks. In contrast, studies of children with autism suggest that such children treat theory-of-mind tasks as logical-reasoning problems, relying primarily on language and other nonsocial cognitive processes in lieu of social insight.

Children with autism generally have executive-function deficits that require planning, flexibility, or working memory combined with inhibitory control (Ozonoff et al., 2004). Performance on false-belief tasks by both typically developing children and children with autism is significantly related to these aspects of executive control (Joseph & Tager-Flusberg, 2004).

Children with better planning skills and inhibitory control are more likely to pass false-belief tasks, supporting the view that successful performance on false-belief tasks requires one to maintain a false representation of an event in working memory while resisting the tendency to predict a person's action on the basis of what one knows to be true.

Language ability has also been closely linked to the development of theory-of-mind skills (see Astington & Baird, 2005). Children with autism who are able to use language to communicate are nevertheless usually delayed in acquiring language and continue to lag behind their peers in basic linguistic abilities. Studies of children with autism report that higher vocabulary-test scores correlate with performance on false-belief tasks (e.g., Happé, 1995). Semantic and grammatical knowledge, as well as more specific knowledge of complex embedded sentence structures, are the most significant predictors of who will pass false-belief tasks (Tager-Flusberg & Joseph, 2005). Children with autism are especially dependent on mastering the syntax and semantics of verbs of communication (for example, "John said that Mary is sleeping") for building theory-of-mind skills. These linguistic constructions provide a format for representing the content of mental states by analogy to the content of speech. For both mental states and speech, the content may differ from reality (e.g., Mary may not be sleeping, even though John said she was, or believed she was). So, through listening and speaking about what people say, some children with autism develop the knowledge that people may represent the world in ways that do not match reality.

Language is important for the development of a consciously mediated explicit theory of mind. About 1 year before they are able to pass false-belief test questions, typically developing toddlers will reliably look to the correct location where Sally will search for the hidden ball (the basket), even while saying that she will look in the box. This eye-gaze behavior is taken as an implicit measure of false-belief understanding, which is the foundation for later-developing explicit knowledge. Ruffman and his colleagues found that, in contrast to well-matched children with mental retardation, children with autism did not look at the correct location on theory-of-mind tasks, although the two groups performed at the same level when answering the verbal test questions (Ruffman, Garnham & Ridout, 2001). Like typically developing toddlers, the children with mental retardation were more likely to pass the implicit measure (eye-gaze) than they were to pass the explicit one (verbal-response), whereas the opposite pattern was found for the children with autism. Ruffman et al. (2001) argue that the implicit measure taps social insight that appears to be lacking in most children with autism, even those who pass false-belief tasks.

Despite the ability of some high-functioning children with autism to pass false-belief tasks, these children still lack social "intuition." Some more able children with autism develop a linguistically mediated theory of mind that provides them with the facility to reason correctly about the social world, but their

theory of mind is not based on the same foundational social insights that are provided by a domain-specific theory-of-mind mechanism. This conclusion is consistent with functional neuroimaging studies, which have shown that high-functioning adults with autism who pass theory-of-mind tasks activate different brain regions when solving such problems. When nonautistic controls process theory-of-mind tasks they typically activate areas in the medial prefrontal cortex and temporo-parietal junction that are considered central to the social-cognitive neural network, as well as areas involved in executive control. In contrast, participants with autism activate only those areas associated with general problem-solving abilities (Frith & Frith, 2003).

THEORY OF MIND AND SYMPTOM SEVERITY IN AUTISM

Surprisingly little research has directly investigated whether core autism symptoms are directly related to theory-of-mind impairments. Early studies reported significant correlations between social or communicative functioning and theory-of-mind performance, but these correlations were no longer significant once age and language level were included as control variables. These negative findings fueled some of the criticisms of the theory-of-mind hypothesis of autism, but these investigations relied on small samples of children who varied widely in age and ability and assessment of theory of mind was limited to false-belief tasks.

More recently, we addressed this issue in a large group of school-aged children with autism using the battery of theory-of-mind tasks developed by Steele et al. (2003). After separating out the effects of age, IQ, and language, theory-of-mind scores were significantly related to scores on the Socialization domain of the Vineland Adaptive Behavior Scales and social- and communication-symptom severity as measured on the Autism Diagnostic Observation Schedule (Tager-Flusberg, 2003). By definition, all children with autism have core deficits in social reciprocity and communication skills that impact their everyday adaptive socialization with peers in school and other community settings. Nevertheless, variation in the severity of these social and communication impairments is partially explained by the degree of impairments in theory of mind.

BEYOND THEORY OF MIND IN AUTISM

Autism involves significant difficulties in understanding mental states. The theory-of-mind hypothesis focuses on deficits in reasoning about mental states. But social and communication developments begin long before theory-of-mind skills emerge in typically developing children. They encompass emotional and perceptual processing that serve as the foundation for social cognition. When the theory-of-mind hypothesis of autism was introduced, Hobson (1993) argued that the deficits associated

with theory of mind were based on early affective impairments. Several research groups have extended studies of mental-state understanding to more naturalistic social contexts to investigate the spontaneous processing of mental-state information from faces, voices, or body gestures (e.g., Klin, Jones, Schultz, & Volkmar, 2003). Thus, the scope of the mechanisms that may underlie the core symptoms of autism has broadened to include on-line perception and responses to a wide range of social stimuli. Even when individuals with autism are able to pass theory-of-mind tasks, they often perform poorly in experiments that tap these core aspects of social/affective information. In the first year of life, social interactions are grounded in recognizing and responding to facial and vocal expression. The earliest signs of autism, including failure to orient toward social stimuli and deficits in joint attention, can be readily interpreted within a broader theory-of-mind framework that encompasses these on-line social-perceptual components as well as more traditional social-cognitive components (Tager-Flusberg, 2001).

FUTURE DIRECTIONS

The theory-of-mind hypothesis stimulated a surge of interest in autism, with a particular focus on how research with this population could provide important insights into the neurocognitive architecture for theory of mind. In recent years, the debate has turned toward the role played by the mirror-neuron system in accounting for the range of social-communicative deficits in autism that include not only those encompassed by theory of mind but also those that go beyond theory of mind—for example, face recognition, imitation, and empathy (Williams et al., 2006). Mirror neurons were first described when researchers noted that certain neurons located in the prefrontal motor cortex of monkeys fired when the animals carried out an action, when they observed the same action carried out by another individual, and when imitating the action, suggesting that these neurons are important for encoding the intentions of other actors (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). Since mirror neurons were initially described, researchers have proposed that the mirror-neuron system is an important component of the social-cognitive network, and there is evidence for impaired functioning of this system in autism (Oberman & Ramachandran, 2007). But there are still many questions about the relationship between the variety of processes that are subsumed under the mirror-neuron umbrella—such as imitation or empathy—and other aspects of social-information processing—including theory of mind and face processing. To investigate the emergence (or nonemergence) of the full range of behaviors that are considered key to the mirror-neuron system and theory of mind, it will be necessary to conduct longitudinal behavioral studies beginning in infancy, before the onset of major autism symptoms. More systematic neuroimaging studies should be conducted to investigate the common and distinct neural pathways that underlie the social-information-processing deficits implicated in autism, with

particular sensitivity to the individual variation that can be expected in this heterogeneous population. The past two decades of research on theory of mind in autism has taught us that no single hypothesis can explain the full range of symptoms that define autism. As we move forward with a broader perspective on the neurocognitive mechanisms that are associated with social-communicative impairments in autism, it will be crucial to embrace the variability that we can expect to find in our data. Only by focusing on this variability can we hope to advance our understanding of this complex and enigmatic neurodevelopmental disorder.

Recommended Reading

- Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D.J. (Eds.). (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.). Oxford, UK: Oxford University Press. A comprehensive sourcebook for developmental, neurobiological, and clinical research studies on theory of mind, including chapters on theory of mind in autism.
- Oberman, L.M., & Ramachandran, V.S. (2007). (See References). A detailed presentation of the mirror-neuron system and its relation to theory of mind in explaining core deficits in autism.
- Ruffman, T., Garnham, W., & Rideout, P. (2001). (See References). An important study that illustrates the difference between implicit and explicit measures of theory-of-mind understanding in children with autism.
- Tager-Flusberg, H. (2001). (See References). A critical analysis and defense of the theory-of-mind hypothesis of autism.

Acknowledgments—Preparation of this article was supported in part by Grant U19 DC 03610 from the National Institute on Deafness and Other Communication Disorders (NIDCD), which is part of the National Institute of Child Health and Human Development/NIDCD-funded Collaborative Programs of Excellence in Autism.

REFERENCES

- Astington, J., & Baird, J. (Eds.). (2005). *Why language matters for theory of mind*. Oxford, UK: Oxford University Press.
- Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a "theory of mind?" *Cognition*, *21*, 37–46.
- Frith, U. (1989). *Autism: Explaining the Enigma*. Oxford, UK: Blackwell.
- Frith, U., & Frith, C. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London – Series B: Biological Sciences*, *258*, 459–473.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.
- 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.
- Hobson, R.P. (1993). *Autism and the development of mind*. Hove, UK: Erlbaum.
- Joseph, R.M., & Tager-Flusberg, H. (2004). The relationship between theory of mind and executive functions to symptom type and

- severity in children with autism. *Development and Psychopathology*, 16, 137–155.
- Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical transactions of the Royal Society of London – Series B: Biological Sciences*, 358, 345–360.
- Oberman, L., & Ramachandran, V. (2007). The simulating social mind: The role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychological Bulletin*, 133, 310–327.
- Ozonoff, S., Cook, I., Coon, H., Dawson, G., Joseph, R.M., Klin, A., et al. (2004). Performance on CANTAB subtests sensitive to frontal lobe function in people with autistic disorder: Evidence from the CPEA Network. *Journal of Autism and Developmental Disorders*, 34, 139–150.
- Peterson, C., Wellman, H., & Liu, D. (2005). Steps in theory of mind development for children with deafness or autism. *Child Development*, 76, 502–517.
- Ruffman, T., Garnham, W., & Rideout, P. (2001). Social understanding in autism: Eye gaze as a measure of core insights. *Journal of Child Psychology and Psychiatry*, 42, 1083–1094.
- 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. (2001). A re-examination of the theory of mind hypothesis of autism. In J. Burack, T. Charman, N. Yirmiya, & P.R. Zelazo (Eds.), *Development in autism: Perspectives from theory and research* (pp. 173–193). Hillsdale NJ: Erlbaum.
- Tager-Flusberg, H. (2003). Exploring the relationships between theory of mind and social-communicative functioning in children with autism. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 197–212). London: Psychology Press.
- Tager-Flusberg, H., & Joseph, R.M. (2005). How language facilitates the acquisition of false belief understanding in children with autism. In J. Astington & J. Baird (Eds.), *Why language matters for theory of mind* (pp. 298–318). Oxford, UK: Oxford University Press.
- Williams, J., Waiter, G., Gilchrist, A., Perrett, D., Murray, A., & Whiten, A. (2006). Neural mechanisms of imitation and 'mirror neuron' functioning in autism spectrum disorders. *Neuropsychologia*, 44, 610–621.