Chapter 3
Intuitive Theory of Mind

3.1 Introducing Theory of Mind

Theory of Mind (ToM) is an umbrella term commonly used to refer to both the commonsense theory and its associated cognitive processes. ToM investigates how people ascribe mental states to other people and how people use mental states to explain and predict the actions of those other persons. ToM explores mindreading, mentalizing or mentalistic abilities, shared by most adults. These abilities are used to treat other agents as ones possessing the unobservable mental or psychological states, actions and processes, which cannot be explicitly perceived. These abilities are also used to anticipate and explain the agents' behavior in terms of such states and processes.

This is how an adult with autism (Wrongplanet 2015) reflects on his ToM capabilities:

> I actually very well remember the time when I considered other people as objects, moving and talking, but devoid of thoughts and feelings. Sometime in my twenties I started looking at people and thinking "Can they possibly have consciousness, same as I do?" This thought seemed preposterous and unworldly. But I finally convinced myself, and now I assume that other people have independents minds and thoughts. This assumption is on the conscious level and disagrees with my intuition.

Two different well known theories have been proposed to explain the basic mechanism underlying the ToM abilities. They are usually referred to as *simulation theory* and *theory-theory* (Vogeley et al. 2001). According to simulation theory, ToM skills are based on taking someone else's view and projecting one's own attitude onto someone else. The simulation approach to reasoning about mental states will be explored in Chap. 5. By contrast, according to theory-theory, theory of mind capacity is based on a distinct body of theoretical knowledge acquired during the individual's ontogenetic development. From the computer science standpoint, ToM is a meta-theory of the theory about mental world. We will investigate what kind of meta-theory is required for the mental world in Sect. 4.1.3.

B. Galitsky, *Computational Autism*, Human–Computer Interaction Series,
DOI 10.1007/978-3-319-39972-0_3

Fig. 3.1 René Descartes's illustration of "Simulation theory" and "theory theory" dualism. Inputs are passed on by the sensory organs to the epiphysis in the brain and from there to the immaterial spirit

Both theory-theory and simulation-theory are actually families of theories. Some theory-theorists maintain that our naïve ToM is a result of our scientific-like exercise of a problem domain capacity to provide a theoretical basis. Other theory-theorists defend a quite different proposal that mindreading relies on the development of a mental organ specifically dedicated to the psychological domain. Simulation-theory also shows different aspects: according to its "moderate" view, mental concepts are not completely excluded from simulation (Fig. 3.1). Simulation can be represented as a procedure through which we:

1. yield and attach to ourselves some mental states of pretense that are intended to correspond to those of the simulated agent;
2. project them onto the target.

By contrast, a stronger version of simulation-based approach denies the superiority of first-person mindreading and proposes that we imaginatively transform ourselves into the simulated agent, interpreting the target's behavior without using any kind of mental concept, not even ones referring to ourselves.

Neurophysiological evidence relevant to theory-theory vs simulation was provided by (Gallese et al. 1996), who demonstrated a mirror neuron system in macaques (Fig. 3.2). Mirror neurons are premotor neurons that are activated when a monkey performs an object-directed action including keeping, capturing, grasping, tearing, manipulating and holding. These neurons are also activated when the animal observes a human experimenter, performing the same class of actions. The discovery of mirror neurons provided a possible mechanism for a simulation theory

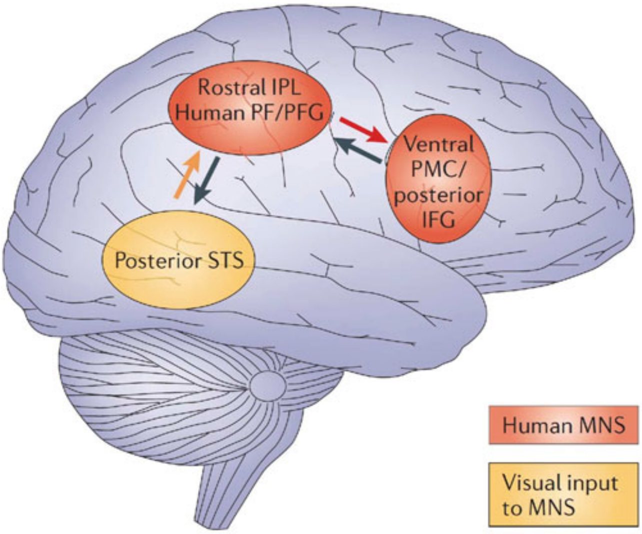

Fig. 3.2 Brain areas involved in the mirror neuron subsystem (Iacoboni and Mazziotta 2007)

account of theory of mind (Gallese and Goldman 1998). Multimodal neurons in motor cortex react to visual observations, helping to understand actions of others by simulating similar motor activity. Distortion in the development of the mirror neural system interferes with the ability to imitate, leading to social impairment and communication difficulties, and may be responsible for the lack of ToM (Iacoboni and Mazziotta 2007).

There has been an important debate in philosophy contrasting 'theory' vs. 'simulation' accounts of reasoning about mental states (see for example (Harris 2000)). This issue is strongly correlated with autism, and in this book we implement a hybrid approach to implementation of ToM engine, merging simulation with implementation of meta-reasoning (Chap. 4). (Stenning 2002) argues that the two may not be as distinct as would at first appear. Chapter 4 of this book is devoted to theory-theory approach, and Chap. 5 describes the simulation approach to implement ToM reasoning.

Before the age of 4, children can impress an external observer that they play together but indeed they play independently, not interacting with each other, each in her own space. Before the age of 3, children do not understand that other children have beliefs, desires and intentions, they live in their own worlds. After the age of 4 children discover that other people have Belief-Desire-Intention (BDI, Rao and Georgeff 1995, Sect. 4.1.2) model of the mental world. Children discover that other people can have wishes not necessarily connected with their own wishes. But initially, before that age children cannot handle this, and they play in parallel worlds. They stop making other people do what they want, but they try to avoid each other and minimize interactions. Then at the age of 5 children start understanding that interaction can work, they start forming beliefs that "under some conditions I can achieve something I want her to do; it is possible to reach an agreement."

3.2 Emphasizing and Systemizing

The empathizing-systemizing theory of autism (Baron-Cohen 2002) proposes that autism spectrum conditions involve deficits in the normal process of empathy, relative to mental age. These deficits can occur by degrees. The notion of empathizing is introduced to cover a broad range of reasoning sub-domains: theory of mind, mind-reading, empathy, and taking the intentional stance (Dennett 1987). We define empathizing as reasoning about the mental world. Empathy includes two elements:

(a) Attribution of mental states and mental actions to oneself and others, as a natural way to make sense of the actions of agents (Baron-Cohen 1994; Premack 1990); and
(b) Emotional reactions that are appropriate in a given mental state.

Since the first test of mind-blindness was administered to children with autism (Baron-Cohen et al. 1985), more than 30 experimental tests have been developed, confirming the impairments in the development of empathizing (Baron-Cohen 1995). The skills of empathizing significantly varies for CwA but are still significantly inferior to that of controls (Fig. 3.3). The limited capabilities in empathizing lead to social and communicative development and in the imagination of others' minds (Baron-Cohen 1987; Leslie 1987).

(Baron-Cohen 2002) attempts to rely on the empathizing-systemizing theory to explain other psychological models such as impairments of executive function or central coherence. From the engineering standpoint, a device can have multiple malfunctions which do not need to be caused by a single subsystem. Nevertheless, in autism research the community attempts to form a single model which would explain the whole range of autistic phenomenology. Even in a reasoning domain, the range of reasoning peculiarities is so broad that it seems hard to find the root cause in the reasoning problems themselves, let along the behavioral autistic features.

Although autism is most often conceptualized as a syndrome of deficits, its altered developmental emphases can also lead to remarkable analytical strengths in some domains. (Baron-Cohen 2002) explains the cognitive superiorities found in autism by the concept of systemizing. It is defined as a drive to analyze objects and events to understand their structure and to predict their future behavior.

Autistic systematizing is based on reduced generalization skills of induction, but fairly efficient rule system once the rules become available to CwA. CwA are good at applying rules to technical systems (such as machines and tools), natural systems (such as biological and geographical phenomena), and abstract systems (such as mathematics or computer programs). Several studies indicate that systemizing in autism is at least in line with mental age, or superior (Baron-Cohen et al. 2003; Lawson et al. 2004). Systemizing may underlie a different set of behavioral features in autism that we refer to as the triad of strengths (Fig. 3.4).

Fig. 3.3 Interaction with horses helps to stimulate empathy

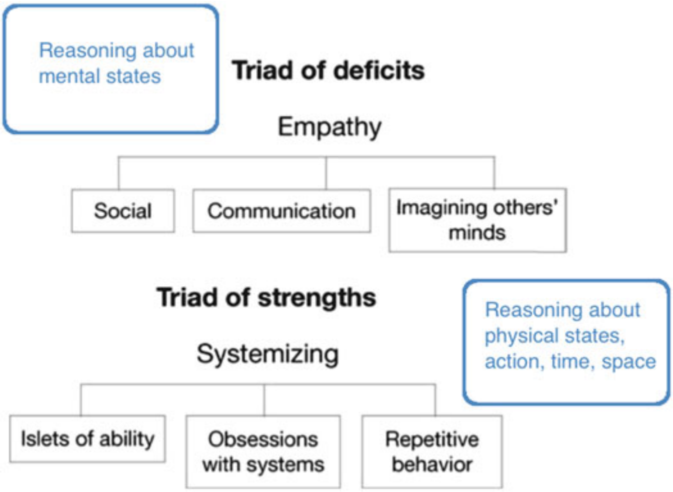

Fig. 3.4 Deficits and strengths of CwA and respective reasoning domains

The outcomes Sally-Ann and 'Smarties' experiments (Sect. 2.2) have been argued to support the 'theory of mind deficit' hypothesis on the cause of autism. Proposed by (Leslie in 1987), it postulates that human beings have evolved a special 'module' devoted specifically to reasoning about other people's minds. As such, this module would provide a cognitive underpinning for empathy. In CC, the module would constitute the difference between humans and their ancestors – indeed, chimpanzees seem to be able to do much less in the way of mind-reading. In CwA, this module would be delayed or impaired, thus explaining abnormalities in communication and also in the acquisition of language, if it is indeed true that the development of joint attention is crucial to language learning (as claimed for instance by Tomasello 1988).

(Norenzayan et al. 2012) found that symptoms of autism correlated with lack of religious belief. They also asked CwA about their empathy (using questions like "I often find it difficult to judge if someone is rude or polite" and "I am good at predicting how someone will feel.").

They found that empathy also correlated with belief. Using a statistical technique of bootstrapping they found that the most plausible explanation for the correlation was that autism was related to a lack of empathy, which in turn was related to lack of belief. In other words, lack of empathy was the 'in between' factor that mediated the relationship between autism and lack of belief. The authors also measured something called systemizing, which is all "about aptitude for, and interest in, reasoning about mechanical and physical objects and processes", and is measure using questions like "I am fascinated by how machines work" and "I find it difficult to understand information the bank sends me on different investment and saving systems". Like empathy, systemizing is correlated both with being male and the degree of autism (although in the opposite direction: autistics are better at systematizing than controls). But, unlike empathy, systematizing does not mediate the effect of autism on religion, in terms of formal correlation.

(Seidner et al. 1988; Stipek and DeCotis 1988) explore memories of emotional experience recounted by high-functioning children with autism and their typically developing peers to mine the depths of children's emotional understanding and discern their strategies for interpreting emotional encounters. Researchers have generated many insights into those types of experiences children consider emotionally evocative by concentrating on the thematic content. This work has shown that high-functioning autistic children demonstrate particularly limited understanding of more complex emotions such as pride, embarrassment, and shame, failing to distinguish these emotions from less complex feelings of similar hedonic tone (e.g., happiness and sadness; Capps et al. 1992).

Yet, without complementary analyses of discourse structure, information on children's strategies for interpreting their emotional experiences is currently lacking. (Losh and Capps 2006) address this problem and consider whether potential differences in discourse structure are restricted to emotional memories or, rather, represent a more pervasive difficulty. They do so by comparing the structural features of children's emotional accounts to those of non-emotional physical states. In view of recent findings of impaired episodic memories in autism (Bowler et al. 2000), the inclusion of non-emotional terms (e.g., sick and tired) is of particular value in the assessment of emotion-specific patterns not assessed in prior studies of autistic individuals' recounted emotional experiences.

3.3 ToM and Other Autistic Accounts

Certain predictions arise, if one considers Empathy-Systemizing and Central Coherence (Sect. 2.5.1) not as mutually exclusive explanations of autistic behavior, but as complementary ones that can be developmentally unified.

Specifically, the attention to detail described by weak central coherence may be one of the earliest manifestations of a strong drive toward systemizing, or vice versa, interest in systemizing may arise as a consequence of attention to detail. As cognitive capacities become more complex and mature, strong "systemizers" may begin to apply some kind of engineering methodology. In this methodology even complex systems are understood by successive local observations in which one input at a time is manipulated while all others are held constant, and effects on the outputs are observed in a similarly sequential manner. This is how an engineering system can be optimized, a causal links between the parameters can be established, or a fault in an engineering system can be discovered. Thus the ultimate effect of the cognitive style described as weak central coherence is not a lack of ability to understand global relationships but rather a difference in the process by which global relationships are established. This is true at least in high-functioning CwA.

Experimental comparison of the ability to make inferences about complex systems, between CwA and controls, and across different stages of development or levels of functioning, may lead to the recognition of Empathy-Systemizing as an elaboration of the Central Coherence model, one that may make more precise and more accurate predictions about the behavior of people with autism when confronted with complex systems. In contrast to controls, CwA use a higher-dimensional representation for learning, so it is more computationally intensive to combine all these dimensions.

Although both central coherence and systemizing are useful psychological models to explain many aspects of autistic behavior, a complete explanation of autism will require that these psychological models be joined with neurobiological substrates—a process complicated by the fact that neither capacity is likely to be atomic in neurobiological terms.

To establish the relations between the psychological theories of autism, logical analysis can be helpful. (Stenning and van Lambalgen 2008) believe that there is a common core to the ToM deficit theory and executive disorder theory, which consists in well-defined failures in non-monotonic reasoning. However, we believe that these deficiencies are very different in nature: non-monotonic reasoning is the logical, domain-independent part, and ToM is a domain-specific set of axioms which happens to be corrupted but can be successfully taught.

3.4 ToM and a Module to Implement It

The notion of a ‘ToM module’ is fairly broad. In the context “from a neural system to behavior” it is obviously meant to be a piece of dedicated neural circuitry. In this way, it can differentiate us from our ancestors and it can also be malfunctioning in isolation. But it is precisely this isolation, (‘encapsulation’ according to Fodor), that is doubtful. One reason is just our general skepticism that evolution does not generally proceed by adding new modules (rather than tweaking old ones), and another is that much of the problem of functionally characterizing human reasoning about minds is about interactions between modules. ToM requires language to formulate beliefs in and it also entails a considerable involvement of working memory, as can be seen in ‘nested’ forms of ToM, as in the example of (Dunbar et al. 2015)

> Shakespeare intended us to realize that Othello believes that Iago knows that Desdemona is in love with Cassio.

Once we understand that it is rather implausible for ToM moduleto operate in isolation, then the ToM deficit hypothesis is becoming less sound. We can now consider the interactions of the ToM module with other language and memory functions, which lead to the possibility that a corruption in these functions is correlated with autism. It is also unclear what the ToM module would have to contain, given the observation that reasoning about intents of others can be partially functional in both CwA and non-human primates.

In this book we differentiate between the general reasoning capabilities and ToM axioms. We believe that they are not interdependent in most occasions. Since we know we can teach ToM axioms successfully, and there is not such axiom that can not be taught to any child, we do not confirm this “modularity” idea.

It is unclear from the experiments at what stage ToM abilities emerge. False-belief tasks were initially proposed as diagnosing a lack of these abilities in normal 3 year-olds and their presence in normal 4-year-olds (Leslie 1987). Others have proposed that irrelevant linguistic demands of these tasks underestimate 3-year-olds’ performance. For example, in the ‘Sally-Anne’ task, the child sees the doll see the sweet placed in one box, and then the child but not the doll sees the sweet moved to another. Now if the child is asked ‘Where will the doll look for the sweet first?’ (instead of ‘Where will the doll look for the sweet?’) then children as young as two can sometimes solve the problem (Siegal and Beattie 1991). This might be read as evidence of the 3-year-olds in the original task adopting a conditional reading of the question (*Where* should the doll look?) rather than a descriptive one (Where will the doll look *first*?). Another possibility associated with a problem in the selection task, is that the younger child’s problem may be with sequencing contingencies in their responses. These arguments push reasoning about intentions earlier in ontogeny.

Hence it is unclear if a neural module for ToM exists. However, it is safe to conclude that ToM-related reasoning belong to a separate clearly circumscribed component detached from the reasoning in other domains such as time, space and other dimensions of physical world.

3.5 ToM in Humans and Animals

Are Theory of Mind abilities unique to humans? (Premack and Woodruff 1978) posed the question: “Does the chimpanzee have a theory of mind”? An affirmative answer would downplay the overall significance of culture and enculturation in human ToM abilities. Reviewing a few decades of experimentation with primates that followed from Premack and Woodruff’s provocative paper, (Call and Tomasello 2008) provide the definitive answer to their question: yes and no. There is solid experimental evidence that chimpanzees understand the goals and intentions of others, as well as the perception and knowledge of others. The behavioral evidence from chimpanzees suggests understanding that goes beyond the reading of surface behaviors of others, to underlying goals and perceptions – at least to the extent that human infants do in similar experimental designs.

In contrast, there is no experimental evidence that chimpanzees can grasp the notion of a false belief, or predict the behavior of another based on what the other knows. If we take a narrow view of the scope of ToM abilities, focusing on social cognitive reasoning, then our closest biological relatives have nothing like our human abilities. If instead we broaden our scope to include social perception and intentional interaction, then chimpanzees are convincingly competent. This shift in research focus toward social competency has led some researchers consider the question for more distant biological relatives, including domesticated dogs and other highly-social animals. The broad set of social skills that are often associated with human’s ToM abilities appear to be common among animals. Birds will hide food far away from potential thieves, and wait to stash food until an onlooker is distracted. Dogs are able to follow a human’s eyes or pointing gestures to hidden food. In contrast, not one other species has passed the false belief test, or exhibited anything like the deep social reasoning which is performed well by humans effortlessly.

ToM starts from (Premack and Woodruff 1978) work on chimpanzees to differentiate between humans and non-human primates. (Leslie 1987) proposed that human beings have a brain ‘module’ that does reasoning about minds, by implementing a ToM, and that autistic reasoning is associated with one or another form of corruption in this module.

So in CC the module constitutes the difference between humans and their ancestors. The work hypothesizes that once chimpanzee acquires ToM, their reasoning would approach humans, and once humans loose parts of ToM, they approach autistic reasoning. At the same time chimpanzees are hyper-social animals, unlike CwA. Whatever cognitive additions yielded humans from their ape ancestors, may be over-represented in autistic cognition. Just for an example to illustrate, much of autistic cognition is an obsessive attempt to extract exception-less truth about a complicated world. This sounds to us rather more like the scientific life than that of chimpanzees. Computer scientists and other natural science academics can empathize with autistic reasoners.

These issues raise many questions concerning what non-human primates are capable of doing in terms of reasoning about behavior and mental processes. Apes are capable of reasoning about the plans of other apes, including the intentions behind their behavior, but they appear not to be able to reason about specific knowledge (epistemic) states. Correspondingly, young children first develop 'desire' psychology before they proceed to 'knowledge' and 'belief' psychology.

3.6 CwA and CC in Abstract Reasoning Tasks

Recent studies (e.g. Dawson et al. 2007) have reported that autistic people perform in the normal range on the Raven Progressive Matrices test, a formal reasoning test that requires integration of relations as well as the ability to deduce behavioral rules and form high-level abstractions. (Morsanyi and Holyoak 2010) compared autistic and control children, matched on age, IQ, and verbal and non-verbal working memory, using both the Raven test and pictorial tests of analogical reasoning. They found that autistic children reasoning capabilities are similar to those of controls on reasoning with relations tests. The authors conclude that the basic ability to reason systematically with relations in the physical world, for both abstract and thematic entities, is intact in autism.

(Gokcen et al. 2009) investigated the potential values of executive function and social cognition deficits in autism. While ToM is generally accepted as a whole, a number of researchers suggested that it can be separated into two components (mental state reasoning and decoding). Both aspects of ToM and verbal working memory abilities were investigated with relatively demanding tasks of mental reasoning for parents of children with autism, who had verbal working memory deficits as well as low performance on a mental state reasoning task. The parents had difficulties in reasoning about others' emotions. In contrast to findings in the control group, low performance of mental state reasoning ability was not associated with working memory deficit in index parents. Social cognition and working memory impairments may represent potential genetic risks associated with autism.

In the physical world, children with autism perform relatively well. Autistic participants outperformed non-autistic participants on abstract spatial tests (Stevenson and Gernsbacher 2013). Non-autistic participants did not outperform autistic participants on any of the three domains (spatial, numerical, and verbal) or at either of the two reasoning levels (concrete and abstract), suggesting similarity in abilities between autistic and non-autistic individuals, with abstract spatial reasoning as an autistic strength.

3.7 ToM Controversy

The term "ToM" is problematic since the "theory" part implies a particular theoretical perspective on how people reason about the "mind". This reasoning happens through the fluid application of theoretical knowledge. The problems with this term have been fruitful since they stimulate psychologists to address the fundamental questions about the role of abstract knowledge (as a classical theoretical construct) in contemporary psychology. "Simulation theory" and "theory theory" dualism can be even considered from the philosophy of mind perspective (Crane and Patterson 2001).

In development psychology, changes in a child's capacity to reason about the mental states of other people has been experimentally observed. One experimental instrument for studying children's abilities to reason about the mental states of others is the False-belief task (Sect. 2.2). Success on this task has been criticized as neither entirely dependent on commonsense psychology abilities nor broadly representative of them (Bloom and German 2000). At the same time, the value of False-belief task is to reliably demonstrating an existence of the developmental shift. (Wellman et al. 2001) aggregated the results of almost two hundred separate studies of the False-belief task, finding that 3-year-olds will consistently fail this task on the majority of trials by indicating that Maxi will look for the object in the location to which his mother has moved it. 4-year-olds will succeed on half the trials, while 5-year-olds will succeed on the majority of trials. (Call and Tomasello 1999) demonstrated that these results are consistent across verbal and non-verbal versions of this task.

There is a developmental change between 3 and 5-year-olds, but it is unclear what exactly is being developed between these ages. One school of thought is that this developmental change can best be characterized as the acquisition by children of a better theoretical model of human psychology, a view first referred to as the "Theory Theory" by philosopher Adam Morton (1980). This view has several advocates among developmental psychologists (Wellman 1990; Gopnik and Wellman 1994), who characterize young children as extremely effective scientists that incrementally adapt their innate knowledge of people to accommodate for their experiences in the world. After years of social interaction, children's developing theories of the mind become more robust in their abilities to predict and explain human behavior, and increasingly include all of the principles of commonsense psychology in Heider's (1958) original characterization. This perspective is consistent with a broader position within developmental psychology that argues that the development of cognitive abilities is best viewed in terms of conceptual change. This perspective follows from the constructivist theories of development advanced by Piaget (1954), and can be contrasted with nativist theories that view the emergence of cognitive abilities as the maturation of innate brain functions.

While robots can acquire some ToM axioms, it is hard to imagine an algorithm that they would learn ToM from their experience. So in terms of reasoning, we hypothesize that all humans have ToM axioms embedded, but CC have this axiom "activated" at the age 4 and CwA are unable to activate it.

(Baron-Cohen et al. 1985) first hypothesized that the main behavioral symptoms of autism could be explained by a deficit in Theory of Mind abilities. The authors compared normal children with those diagnosed with autism and Down's syndrome on a variant of false-belief task involving two dolls, Sally and Anne. Even though the mental age of the autistic children was higher than that of the other groups, they alone failed to correctly ascribe a false belief to the doll in the experiment. The finding sparked a vigorous theoretical debate among the community of developmental psychologists and autism researchers that continues today.

Tager-Flusberg (2007) reflects on two decades of research that followed Baron-Cohen et al.'s hypothesis, which has upheld the original result: children with autism have difficulty attributing mental states to themselves or to other people. However, the significance of this finding is in doubt. Deficits in ToM abilities are not universal among autistic children, and neither offer an explanation for other typical symptoms such as repetitive nor for restricted behavior patterns. Tager-Flusberg advises to avoid a narrow view of the social-cognitive deficits in autism, and refers to recent studies on children's perception of mental-state information in faces, voices, and body gestures (Grigorenko et al. 2003). If the connection between ToM abilities and autism is to be explanatory, then the traditional understanding of ToM must be broadened to includes these social-perceptual skills. In today's corpus of work, the relationship between ToM and autism is fairly complex to serve as an illustration for the nativist-constructivist debate.

3.8 Discussion and Conclusions

People with autism and machines sometimes have difficulty comprehending when other people and users of these machines do not know something. CwA can get very agitated when a peer does not know the answer to a question she asks. By not understanding that other people think, believe, know, and want differently than themselves, CwA have problems relating socially and communicating to other people. As a result CwA and computer systems are frequently unable to anticipate what others will say or do in various situations, and have difficulty understanding that their classmates even have thoughts and emotions.

ToM arose from the study of primates and their social organization, and scholars in many fields – philosophy, anthropology, psychology, psychiatry and neuroscience – have contributed to this expanding topic.

For the purpose of a better representation and treatment, a more concise, more formal representation for reasoning about mental world than ToM is required. Since CwA have strong systematizing skills, they should be the foundation to ground the emphasizing skills. Since they cannot be introduced in a natural way, they should be taught via rules, such as an empathy to someone's pain, a scenario to pretend, a knowledge state to ask or share information.

To enable ToM to better correlate reasoning about mental attitudes and behavior, the notion of knowing about knowing needs to be formalized and expressed as axioms. Moreover, relations between knowledge space and intention space needs to be established as a rule system suitable for teaching CwA and machines (Chap. 4). A formal link between ToM as a theory-theory and meta-reasoning (Sect. 4.1.3) needs to be established. We need to marry ToM which ascribes mental states to humans with the multiagent systems theory which accumulated substantial experience doing this for automated agents: it will happen in Chaps. 4 and 5.

The foundation of ToM are connected with the nativist-constructivist debate which has been initiated by philosophers hundreds of years ago. It concerns the origin of knowledge and whether it is yielded by native abilities or was derived empirically. In terms of linguistic capabilities, this is formulated as whether humans possess a specific cognitive mechanism for comprehending and producing language, or these capabilities are due to a general cognitive tools. The former represents the nativists theories and the latter is favored by constructivists. If an individual has an innate grammatical knowledge, it has to be domain-specific. Also, a deviation, a move away from this grammatical knowledge means that a language is not associated with special cognitive skills.

Computer science favors the nativist positions and robots need separate components for each kind of knowledge. Teaching CwA, however, we intend to give them general axioms about knowledge and then expect these axioms to be applied in multiple modalities beyond language. How can we teach children to classify states and words for them into abstract categories, unless they already have knowledge of these categories? To overcome this difficulty, we will introduce a basis of undefined concepts and then teach other concepts of mental world relying on this basis.

If robots are capable to recognize faces, voices and body gestures and have functioning ToM components, it can be shown that removal of some axioms in the ToM component will break the overall system, having signal recognition component intact (Galitsky 2002).

As to our framework of *reasoning engine* → *behavior*, we now focus on the main component of reasoning such as reasoning about mental states as the cornerstone of autistic reasoning. We target explaining the broad range of autistic behavior via the peculiarities of autistic reasoning in this very restricted domain, putting the whole physical world aside. We will look at emphasizing from the systemizing standpoint and attempt to represent the richness of mental world in a formal, structured way acceptable to an autistic systematizer.

References

Baron-Cohen S (1987) Perception in autistic children. In: Cohen D (ed) Handbook of autism and pervasive developmental disorders. Wiley, Hoboken

Baron-Cohen S (1995) Mindblindness: an essay on autism and theory of mind. MIT Press/Bradford Books, Boston

Baron-Cohen S (2002) The extreme male brain theory of autism. Trends Cogn Sci 6:248–254

Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a 'theory of mind'? Cognition 21:37–46
Baron-Cohen S, Richler J, Bisarya D, Gurunathan N, Wheelwright S (2003) The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. Philos Trans Biol Sci 358:361–374
Bloom P, German TP (2000) Two reasons to abandon the false belief task as a test of theory of mind. Cognition 77(2000):B25–B31
Bowler D, Gardner J, Grice S (2000) Episodic memory and remembering in adults with Asperger syndrome. J Autism Dev Disord 30:295–304
Call J, Tomasello M (1999) A nonverbal false belief task: the performance of children and great apes. Child Dev 70(2):381–95
Call J, Tomasello M (2008) Does the chimpanzee have a theory of mind? 30 years later. Trends Cogn Sci 12(5):187–192
Capps L, Yirmiya N, Sigman M (1992) Understanding of simple and complex emotions in non-retarded children with autism. J Child Psychol Psychiatry 33:1169–1182
Cohen IL (1994) An artificial neural network analogue of learning in autism. Biol Psychiatry 36:5–20
Crane T, Patterson S (2001) "Introduction". History of the Mind-Body Problem. pp. 1–2. the assumption that mind and body are distinct (essentially, dualism)
Dawson M, Soulires I, Gernsbacher MA, Mottron L (2007) The level and nature of autistic intelligence. Psychol Sci 18(8):657–662
Dennett D (1987) The intentional stance. MIT Press, Cambridge, MA
Dunbar RIM, Launay J, Curry O (2015) The complexity of jokes is limited by cognitive constraints on mentalizing. Hum Nat 1–11
Galitsky B (2002) Extending the BDI model to accelerate the mental development of autistic patients. In: Second international conference on development & learning. Cambridge, MA
Gallese V, Goldman A (1998) Mirror neurons and the simulation theory of mind-reading. Trends Cogn Sci 2:493–501
Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. Brain 119:593–609
Gokcen S, Bora E, Erermis S, Kesikci H, Aydin C (2009) Theory of mind and verbal working memory deficits in parents of autistic children. Psychiatry Res 166(1):46–53
Gopnik A, Wellman HM (1994) The theory theory. In: Hirschfeld LA, Gelman SA (eds) Mapping the mind: domain specificity in cognition and culture. Cambridge University Press, New York
Grigorenko EL, Klin A, Volkmar F (2003) Hyperlexia: disability or superability? J Child Psychol Psychiatry 44:1079–1091
Harris PL (2000) The work of the imagination. Blackwell, Oxford/Boston
Heider F (1958) The psychology of interpersonal relations. Wiley, New York
Iacoboni M, Mazziotta JC (2007) Mirror neuron system: basic findings and clinical applications. Ann Neurol 62:213–218
Lawson J, Baron-Cohen S, Wheelwright S (2004) Empathising and systemising in adults with and without Asperger syndrome. J Autism Dev Disord 34:301–310
Leslie AM (1987) Pretence and representation: the origins of "theory of mind". Psychol Rev 94:412–426
Losh M, Capps L (2006) Understanding of emotional experience in Autism: insights from the personal accounts of high-functioning children with Autism. Dev Psychol 42(5):809–818
Morsanyi K, Holyoak KJ (2010) Analogical reasoning ability in autistic and typically developing children. Dev Sci 13(4):578–587
Morton A (1980) Frames of mind. Oxford University Press, Oxford
Norenzayan A, Gervais WM, Trzesniewski KH (2012) Mentalizing deficits constrain belief in a personal god. PlosOne http://dx.doi.org/10.1371/journal.pone.0036880
Piaget J (1954) The construction of reality in the child. Basic Books, New York
Premack D (1990) The infant's theory of self-propelled objects. Cognition 36:1–16

Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? Behav Brain Sci 4:515–526

Rao AS, Georgeff MP (1995) BDI-agents: from theory to practice. In: Proceedings of the first international conference on multiagent systems (ICMAS' 1995)

Seidner LB, Stipek D, Feshbach N (1988) A developmental analysis of elementary school-aged children's concepts of pride and embarrassment. Child Dev 59:367–377

Siegal M, Beattie K (1991) Where to look first for children's knowledge of false beliefs. Cognition 38:1–12

Stenning K (2002) Seeing reason. Image and language in learning to think. Oxford University Press, Oxford

Stenning K, Michiel Van L (2008) Human reasoning and cognitive science. MIT Press, Cambridge, MA

Stevenson JL, Gernsbacher MA (2013) Abstract spatial reasoning as an Autistic strength. PlosOne. doi:10.1371/journal.pone.0059329

Stipek D, DeCotis K (1988) Children's understanding of the impli- cations of causal attributions for emotional experiences. Child Dev 59:1601–1610

Tager-Flusberg H (2007) Evaluating the theory-of-mind hypothesis of autism current directions. Psychol Sci 16(6):311–315

Tomasello M (1988) The role of joint-attentional processes in early language acquisition. Lang Sci 10:69–88

Vogeley K, Bussfeld P, Newen A, Herrmann S, Happé F, Falkai P, Maier W, Shah NJ, Fink GR, Zilles K (2001) Mind reading: neural mechanisms of theory of mind and self-perspective. Neuroimage 14(1 Pt 1):170–181

Wellman HM (1990) The child's theory of mind. MIT Press, Cambridge, MA

Wellman HM, Cross D, Watson J (2001) Meta-analysis of theory-of-mind development: the truth about false belief. Child Dev 72:655–684

Wrongplanet (2015) Hide and seek (And Mindblindness). http://wrongplanet.net/forums/viewtopic.php?t=738